KFKI 1979-82

COLLECTION OF SCIENTIFIC PAPERS
IN COLLABORATION WITH JOINT INSTITUTE FOR
NUCLEAR RESEARCH, DUBNA, USSR AND
CENTRAL RESEARCH INSTITUTE FOR PHYSICS,
BUDAPEST, HUNGARY

ALGORITHMS AND PROGRAMS FOR SOLUTION OF SOME PROBLEMS IN PHYSICS

THIRD VOLUME

Hungarian Academy of Sciences

CENTRAL
RESEARCH
INSTITUTE FOR
PHYSICS

BUDAPEST

KFKI-1979-82

СОВМЕСТНЫЙ НАУЧНЫЙ СБОРНИК ОБЬЕДИНЕННОГО ИНСТИТУТА ЯДЕРНЫХ ИССЛЕДОВАНИЙ /ДУБНА, СССР/ И ЦЕНТРАЛЬНОГО ИНСТИТУТА ФИЗИЧЕСКИХ ИССЛЕДОВАНИЙ /БУДАПЕШТ, ВЕНГРИЯ/

Алгоритмы и программы для решения некоторых задач физики

Выпуск третий

Ответственный за выпуск: Е.П. Жидков

Редакторы: Г. Немет̆х, Б. Н. Хоромский

COLLECTION OF SCIENTIFIC PAPERS IN COLLABORATION WITH JOINT INSTITUTE FOR NUCLEAR RESEARCH, DUBNA, USSR AND CENTRAL RESEARCH INSTITUTE FOR PHYSICS, BUDAPEST, HUNGARY

Algorithms and Programs for Solution of Some Problems in Physics

Third volume

Responsible person in edition: E.P. Zhidkov

Editors: G. Németh, B.N Horomskii

HU ISSN 0368-5330

ISBN 963 371 608 X

С О Д Е Р Ж А Н И Е

C O N T E N T

ИССЛЕДОВАНИЕ СЛОЖНЫХ ПЕРОВСКИТНЫХ СОЕДИНЕНИЙ МЕТОДОМ МЕССБАУЭРОВСКОЙ СПЕКТРОСКОПИИ

А.Н.Тихонов, В.Я.Галкин, В.П.Горьков, Р.Н.Кузьмин
Х.Х.Ройг Нуньес, В.В.Шагдаров

Московский государственный университет
им. М.В. Ломоносова, Москва

АННОТАЦИЯ

В работе изучаются сложные перовскитные соединения с частичным упорядочением В-ионов. Построены математические модели мессбауэровских спектров этих соединений, проведен анализ зависимости модельных спектров от некоторых физических характеристик /решены прямые задачи/, построены оценки параметров моделей и установлено их соответствие экспериментальным спектрам /для чего решены соответствующие обратные задачи/. Это позволило изучить распределения В-ионов в кристаллической решетке и связать их с температурой магнитного упорядочения сложных перовскитных соединений.

ABSTRACT

In the present paper the compound perovsceet combinations with the partially regulated B-ions are studied. Mathematical models of Mössbauer spectra of these combinations are constructed and dependence of the model spectra on some physical characteristics is analysed. The parameters of the models are estimated and their correspondence to the experimental spectra is ascertained. It makes possible study of the distributions of B-ions in the crystalline lattice and gives their connection with the temperature of magnetic regulation of the compound perovsceet combinations.

Окислы со структурой перовскита вызывают особый интерес, обусловленный разнообразием их электрических и магнитных свойств. В последние годы пристальное внимание уделяется окислам сложных составов с химической формулой $AB'_x B''_{1-x} O_3$. В указанных соединениях ионы и располагаются в центрах кислородных октаэдров, а их распределения оказывают сильное влияние на физические свойства перовскитов. До сих пор использование мессбауэровской спектроскопии для изучения распределения B-ионов ограничивалось наиболее простыми случаями полностью неупорядоченных или - упорядоченных перовскитных соединений [I,2] . Существенно более сложными объектами для изучения являются перовскитные соединения с частичным упорядочением B-ионов. Предметом нашего исследования были сложные перовскитные соединения $SrFe_{2/3}W_{1/3}O_3$, $CaFe_{2/3}W_{1/3}O_3$ и $CdFe_{1/2}Nb_{1/2}O_3$.

Целями настоящей работы являлись построение математических моделей мессбауэровских спектров исследуемых соединений, анализ зависимости модельных спектров от некоторых физических параметров (решение прямых задач), оценка параметров моделей и их соответствие экспериментальному спектру (обратные задачи). Это позволило изучить распределения B-ионов в кристаллической решетке и связать их с температурой магнитного упорядочения сложных перовскитных соединений.

Для описания математической модели (теоретического представления мессбауэровского спектра сложных перовскитных соединений) будем учитывать влияние на мессбауэровское ядро ионов в ближайших B-позициях. Возможны десять вариантов ($j = 1,2,\ldots,10$) взаимного расположения ионов B' и B'', сведенные в таблицу I. Каждому значению j соответствует пара чисел n и k ,где n -число ионов B' в ближайшем окружении,

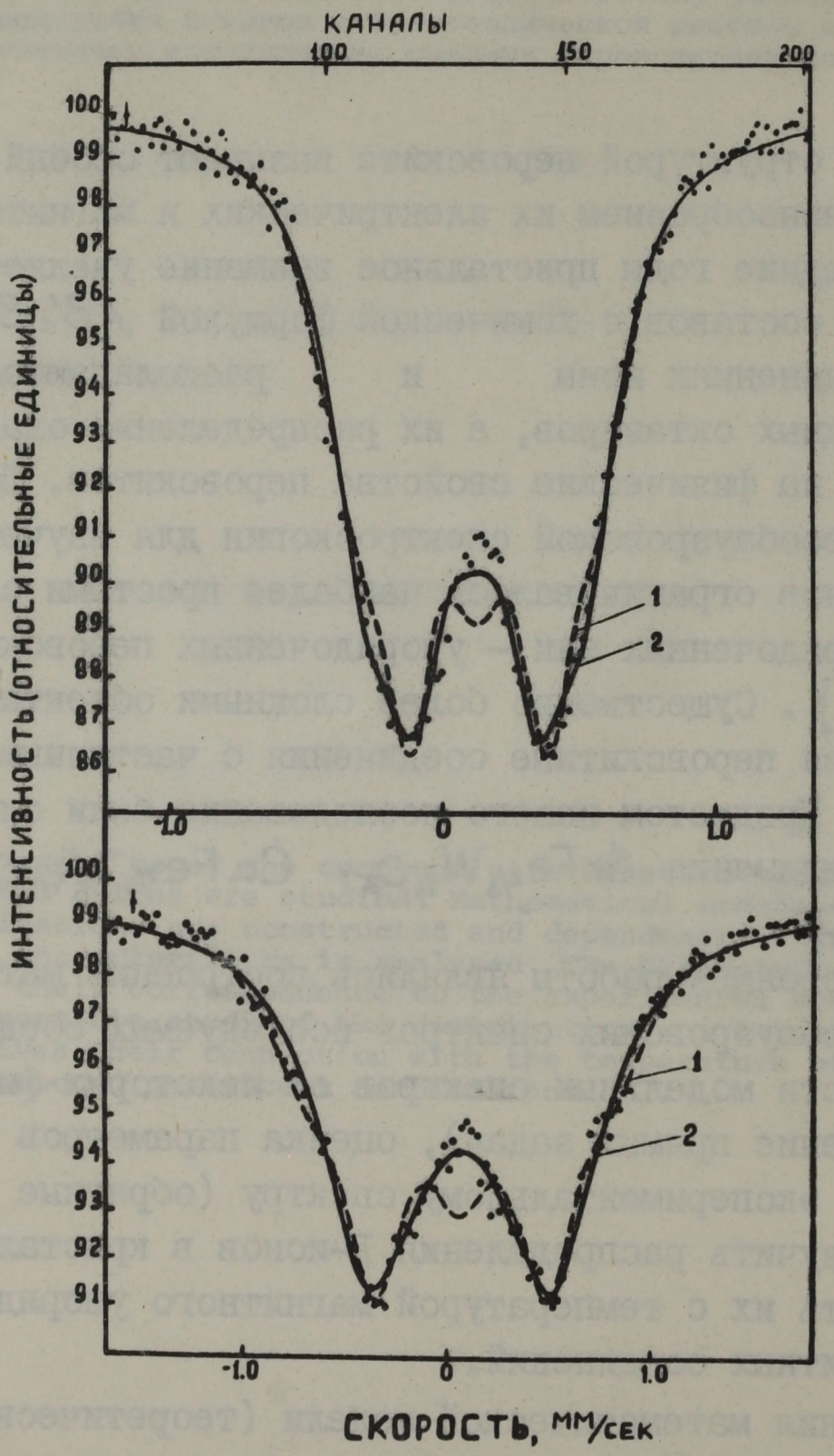

КАНАЛЫ
100
150
200
ИНТЕНСИВНОСТЬ (ОТНОСИТЕЛЬНЫЕ ЕДИНИЦЫ)
100
99
98
97
96
95
94
93
92
91
90
89
88
87
86
-1.0
0
1.0
1
2
СКОРОСТЬ, мм/сек

Рис.I.

которые создают квадрупольное расщепление $\Delta E^j = K \cdot \Delta E$. В работах [I,3] показано, что ионы B' и B'' образуют группы, приводящие к четырем значениям квадрупольного расщепления: $K = 0, I, \sqrt{3}, 2$ (см.табл.I).

Общее представление для математического ожидания экспериментального мессбауэровского спектра в предположении, что каждое возможное окружение B' дает вклад в спектр, имеет вид [3]

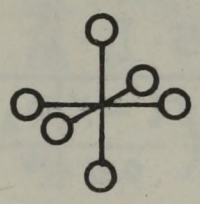
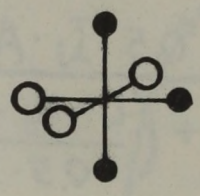
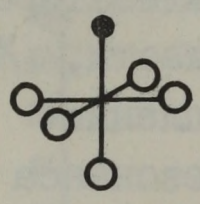
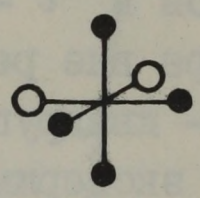
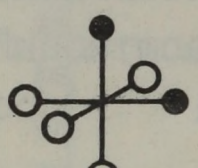
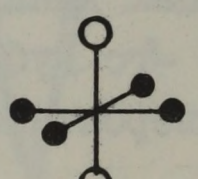
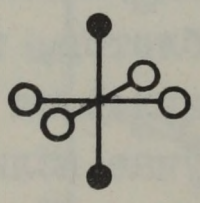
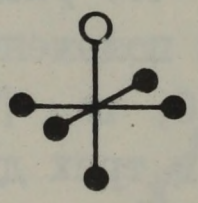
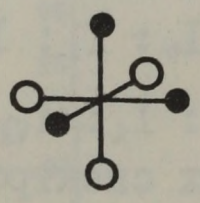
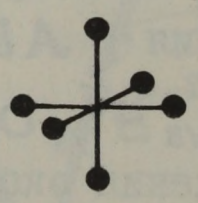
$$MN_i = N_\infty - \sum_{j=1}^{10} \left[\frac{0.5 \cdot I_j \cdot A}{1 + \left(\frac{i - X_j - \Delta E^j}{0.5 \cdot \Gamma} \right)^2} + \frac{0.5 \cdot I_j \cdot A}{1 + \left(\frac{i - X_j + \Delta E^j}{0.5 \cdot \Gamma} \right)^2} \right] \quad (I)$$

где N_i – набор импульсов в i –ом канале анализатора, N_∞ – представление набора вне резонансной области, X_j –центр j –го дублета, ΔE^j – квадрупольное расщепление j –го дублета, Γ – ширина экспериментального резонанса (X_j , ΔE^j , Γ – в единицах канала), I_j – интенсивность резонансного поглощения j –го дублета, A – масштабный множитель, I_j совпадают с вероятностями $P(n)_K$ соответствующих окружений B' .

Если предположить, что различные окружения мессбауэровского ядра не влияют на положения центров дублетов, т.е. в (I) $X_j = X$, $j = I,2,\ldots,I0$, то MN_i представляется суперпозицией одиночной линии и трех дублетов. В случае полностью неупорядоченного состояния получим соотношение интенсивностей этих линий для соединения $AB'_{1/2}B''_{1/2}O_3$ $I_0 : I_1 : I_{\sqrt{3}} : I_2 =$ 5:I8:6:3, а для $AB'_{2/3}B''_{1/3}O_3$ – $I_0 : I_1 : I_{\sqrt{3}} : I_2 =$ = 43:I48:32:20. Приближения экспериментальных спектров соединений $CdFe_{1/2}Nb_{1/2}O_3$ и $SrFe_{2/3}W_{1/3}O_3$ в рассматриваемой модели приведены на рис.I (кривые I) и являются явно неудовлетворительными. Возникает вопрос об упорядочении ионов B' и B'' в кристаллической решетке.

Перейдем к расчету вероятностей окружений в случае упорядоченного состояния. Рассмотрим распределение B–ионов в кристаллической решетке, узлы которой представляют центры кислородных октаэдров. Упорядоченное состояние характеризуется размещением различных B–ионов по неэквивалентным узлам такой решетки. Пусть в решетке существует r различных типов узлов с концентрациями ν_ℓ , $\ell = I,2,\ldots, r$ для каждого типа.

Таблица I.

	j	n	κ		j	n	κ
	1	0	0		6	3	$\sqrt{3}$
	2	1	1		7	4	1
	3	2	1		8	4	2
	4	2	2		9	5	1
	5	3	0		10	6	0

● B' ○ B''

Назовем нахождение ионов B' и B'' в узлах типа ℓ событиями A'_ℓ и A''_ℓ. Тогда для вероятностей $p'_\ell = P\{A'_\ell\}$, $p''_\ell = P\{A''_\ell\}$ в предположении, что решетка не содержит вакансий, имеем $p'_\ell + p''_\ell = I$ при любом ℓ. Для каждого сорта ионов выполняются условия [4]

$$\sum_{\ell=1}^{r} \nu_\ell \cdot p'_\ell = c', \qquad \sum_{\ell=1}^{r} \nu_\ell \cdot p''_\ell = c'', \tag{2}$$

где c', c'' – концентрации B', B'' соответственно.

Пусть r' типов узлов предназначены для ионов B', а $r'' = r - r'$ типов узлов – для ионов B''. Если предположить, что вероятности нахождения ионов одного сорта в "своих" узлах не зависят от типа узлов, то

$$p'_1 = p'_2 = \cdots = p'_{r'} = p, \quad p''_{r'+1} = p''_{r'+2} = \cdots = p''_r = q, \tag{3}$$

где p и q означают вероятности нахождения ионов B' и B'' в "своих" узлах. При подстановке (3) в (2) имеем

$$p \cdot \sum_{\ell=1}^{r'} \nu_\ell + (1-q) \cdot \sum_{\ell=r'+1}^{r} \nu_\ell = c' \tag{4}$$

Введем параметр дальнего порядка S следующим образом

$$S = \frac{p - c'}{1 - c'}. \tag{5}$$

Из (4) и (5) имеем следующие зависимости вероятностей от S

$$p = c' + (1-c') \cdot S, \quad q = (1-c') \cdot \left[1 + \frac{\sum_{\ell=1}^{r'} \nu_\ell}{\sum_{\ell=r'+1}^{r} \nu_\ell} S\right] \tag{6}$$

Вероятность $P(n)_к$ можно записать в виде

$$P(n)_k = \sum_{\ell=1}^{r} C'_\ell \cdot P_\ell(n)_k , \qquad (7)$$

где $C'_\ell = \frac{\nu_\ell \cdot P'_\ell}{c'}$ – концентрация (доля) ионов B' в ℓ –ом узле, $P_\ell(n)_k$ – вероятность нахождения n ионов B' в ближайшем окружении узла ℓ –го типа, создающих квадрупольное расщепление ΔE^j .

Для того, чтобы определить вероятности $P_\ell(n)_k$. рассмотрим узел ℓ –го типа и обозначим узлы его ближайшего окружения через ℓ_s , $\ell = 1,2,\ldots, r$, s – порядковый номер соседнего узла, $s = 1,2,\ldots, z$, z – координационное число. Назовем событиями $A_{\ell_s}^{\beta_s}$, $\beta_s = 1,2$ нахождение ионов B' и B'' соответственно в узле ℓ_s и обозначим $P_{\ell_s}^{\beta_s} = P\{ A_{\ell_s}^{\beta_s} \}$. Одновременное расположение ионов любого сорта в узлах $\ell_1,\ldots, \ell_z$ есть событие

$$A\begin{pmatrix} \beta_1,\ldots,\beta_z \\ \ell_1,\ldots,\ell_z \end{pmatrix} = \bigcap_{s=1}^{z} A_{\ell_s}^{\beta_s} . \qquad (8)$$

Согласно теории упорядочения [5,6] , в которой предполагается отсутствие корреляций при расположении ионов,

$$P\left\{A\begin{pmatrix} \beta_1,\ldots,\beta_z \\ \ell_1,\ldots,\ell_z \end{pmatrix}\right\} = \prod_{s=1}^{z} P_{\ell_s}^{\beta_s} . \qquad (9)$$

При числе n ионов B' в ближайшем окружении существует C_n^z способов расположений, каждый из которых описывается совокупностью $\{\beta_s\} = (\beta_1,\ldots,\beta_z)$. В общем случае различные совокупности $\{\beta_s\}$ при фиксированном n приводят к разным значениям ΔE^j . Из общего множества совокупностей $\{\beta_s\}$ можно выделить подмножество C^k совокупностей $\{\beta_s\}^k$, дающих одно и то же значение вкадрупольного расщепления $k\cdot\Delta E$. Для определенного типа ℓ –го узла каждая совокупность $\{\beta_s\}^k$ определяет событие $A\begin{pmatrix} \beta_1,\ldots,\beta_z \\ \ell_1,\ldots,\ell_z \end{pmatrix}^k$ вероятность которого дается формулой (9). Поскольку все события $A\begin{pmatrix} \beta_1,\ldots,\beta_z \\ \ell_1,\ldots,\ell_z \end{pmatrix}^k$ могут иметь

место в кристалле, то появление в спектре дублета с квадрупольным расщеплением $k \cdot \Delta E$, обусловленное n ионами B' в ближайшем окружении узла ℓ-го типа, определяется событием

$$\mathcal{A}_\ell(n)_k = \bigcup \mathcal{A}\left(\begin{matrix}\beta_1, \ldots, \beta_z \\ \ell_1, \ldots, \ell_z\end{matrix}\right)^k, \qquad (10)$$

где объединение проводится по всем элементам подмножества C^k. Вследствие несовместимости указанных событий при фиксированном ℓ имеем

$$P_\ell(n)_k = P\{\mathcal{A}_\ell(n)_k\} = \sum P\left\{\mathcal{A}\left(\begin{matrix}\beta_1, \ldots, \beta_z \\ \ell_1, \ldots, \ell_z\end{matrix}\right)^k\right\}, \qquad (11)$$

где суммирование выполняется по тем же элементам.

Подмножество C^k в общем случае можно разбить на u классов $C_1^k, \ldots, C_u^k$ таких, что вероятности соответствующих каждому классу событий одинаковы. Тогда

$$P_\ell(n)_k = \sum_{\mu=1}^{u} f_\mu \cdot P_\mu, \qquad (12)$$

где P_μ – вероятность события, вошедшего в класс C^k (P_μ определяется в (9)), f_μ – число элементов класса C^k и равняется индексу точечной подгруппы симметрии $H\left(\begin{matrix}\beta_1, \ldots, \beta_z \\ \ell_1, \ldots, \ell_z\end{matrix}\right)^k$ расположения ионов B' и B'' в ближайшем окружении узла ℓ-го типа относительно точечной группы симметрии $G(\ell_1, \ldots, \ell_z)$ окружения узла ℓ-го типа узлами типа $\ell_1, \ldots, \ell_z$. Если m –порядок группы $G(\ell_1, \ldots, \ell_z)$ и m_μ – порядок подгруппы $H\left(\begin{matrix}\beta_1, \ldots, \beta_z \\ \ell_1, \ldots, \ell_z\end{matrix}\right)$, тогда

$$f_\mu = \frac{m}{m_\mu}. \qquad (13)$$

Таким образом, соотношения (13), (12) и (7) определяют

искомые значения вероятностей $P(n)_k$.

Для исследуемых соединений был предложен ряд конкретных моделей упорядоченных структур. Для этих моделей можно выписать выражения вероятностей $P(n)_k$ через параметр S и рассчитать их. Такие расчеты и графические зависимости вероятностей от S приведены в [3].

Дальнейшее уточнение математической модели мессбауэровского спектра связано с посылками на параметры X_j . Исследованы три варианта влияния окружений на положения центров дублетов, которые характеризуют изомерные сдвиги. В первом варианте считается, что положения центров дублетов не зависят от различных окружений мессбауэровских ядер: $X_j = X$. Второй вариант заключается в том, что разным числам n ионов B' в окружении соответствуют свои центры дублетов. Третий вариант отвечает предположению, что разность d между положениями центров дублетов мессбауэровских ядер, окруженных n ионами B' и нулем ионов B' , пропорциональна числу ионов соответствующего окружения $d = n \cdot \Delta$, где Δ - коэффициент пропорциональности.

Выбрав модель структуры решетки, мы конкретизируем в представлении (I) параметры I_j , ΔE^j , X_j , j = I,2,..., I0. Параметрическое представление математического ожидания и связи на I_j , ΔE^j , X_j определяют математическую модель мессбауэровского спектра. Поскольку об исследуемом физическом явлении возможно несколько теоретических посылок, это всякий раз приводит к соответствующей модели мессбауэровского спектра.

Сам спектр для выбранной упорядоченной структуры полностью определен, если заданы значения параметров S , ΔE^j , X_j , а также параметра Γ -ширины резонансных линий, параметра фона N_∞ и масштабного коэффициента A (отношение $A : N_\infty$ представляет собой величину эффекта Мессбауэра). Задаваясь реальными значениями перечисленных параметров, мы можем решить прямую задачу: рассчитать модельный мессбауэровский спектр и исследовать зависимость его формы от интересующих нас параметров. Без такого исследования невозможна разумная постановка обратной задачи - интерпретации экспериментального мессбауэровского спектра: оценка искомых параметров в представлении (I) по результатам эксперимента или проверка

соответствия физических моделей изучаемому экспериментальному спектру. Это объясняется тем, что обратная задача может оказаться вырожденной по некоторым параметрам или неустойчивой к входным данным.

Для соединения $SrFe_{2/3}W_{1/3}O_3$ с решеткой, в которой часть ионов $B'(Fe)$ образуют цепочки, а другая часть ионов B'' образует цепочки с ионами B' (B' чередуются через три иона B'') и первом варианте посылки об изомерном сдвиге зависимость формы модельного спектра от S приведена на рис. 2. Видно, что при малых S ($S \leq 0.3$) существенен вклад одиночной линии (наличие минимума в центральной части спектра). С увеличением S интенсивность дублета с максимальным расщеплением монотонно возрастает, а интенсивности остальных линий уменьшаются. Это приводит к тому, что модельный спектр с ростом S расширяется, хотя параметры ΔE и Γ фиксированы.

Проведен также анализ зависимости формы модельных спектров от величины изомерного сдвига. Результаты исследования указанных прямых задач подробно изложены в [3]. Оказывается, для некоторых упорядоченных структур решеток существуют различные наборы S , ΔE , Γ , которые дают модельные спектры, близкие по форме.

Интерпретация экспериментального мессбауэровского спектра формализуется как некоторая обратная математическая задача. Первый аспект интерпретации при фиксированной модели структуры решетки заключается в оценивании искомых физических величин по косвенным измерениям N_i . Второй аспект интерпретации в рассматриваемом случае состоит в проверке соответствия каждой из посылок об изомерном сдвиге результатам эксперимента.

Экспериментальный мессбауэровский спектр трактуется как N -мерный случайный вектор N_i , $i = 1,2,\ldots, N$, с независимыми пуассоновскими компонентами, математическое ожидание которого задается соотношением (I). Как только фиксируются модели структуры решетки и изомерного сдвига, конкретизируется представление (I), т.е. математическая модель $MN_i = MN_i(\alpha)$, где α - вектор параметров размерности m , который определяет модель.

Для оценки компонент вектора α был использован метод

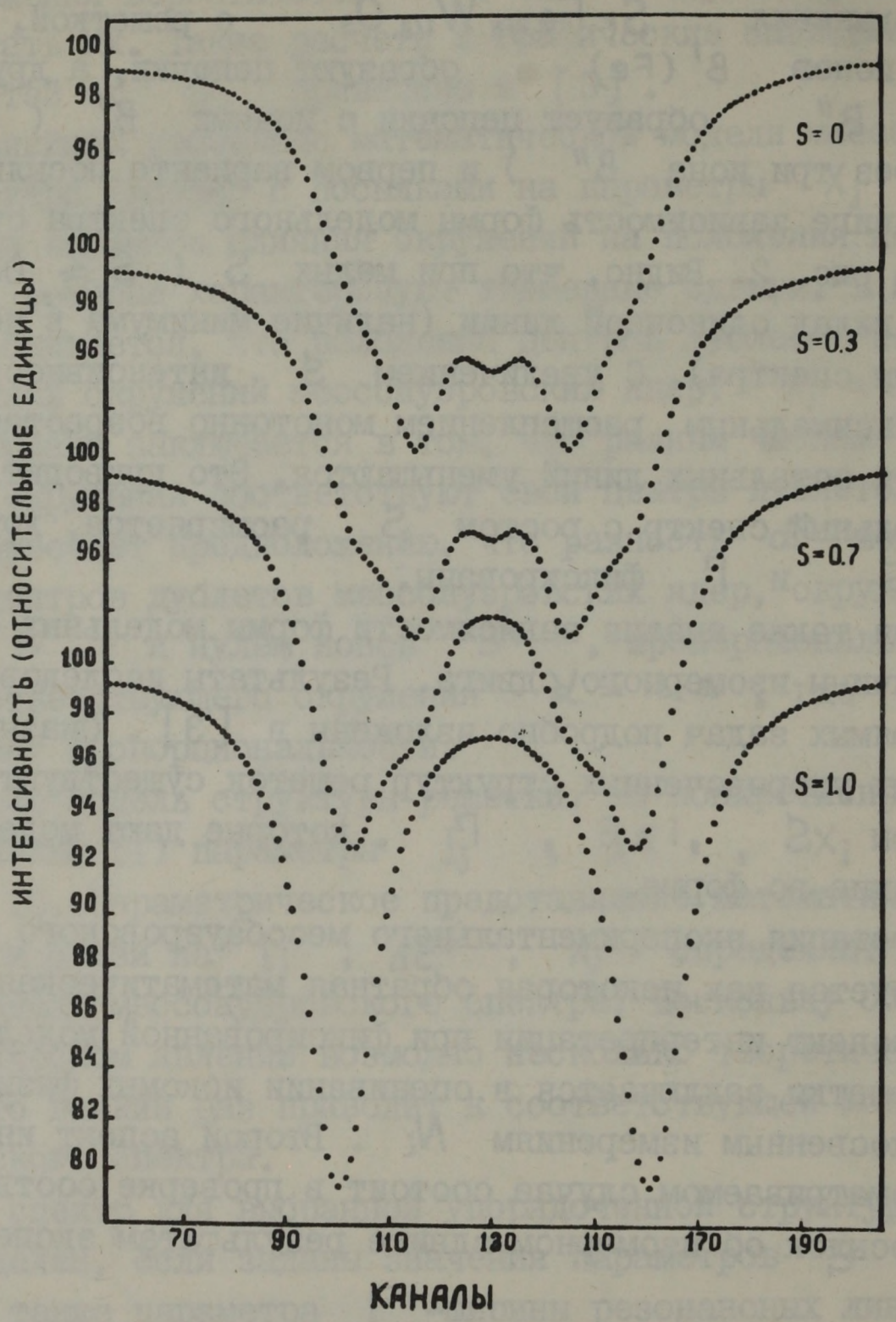

ИНТЕНСИВНОСТЬ (ОТНОСИТЕЛЬНЫЕ ЕДИНИЦЫ)
100
98
96
100
98
96
100
98
96
100
98
96
94
92
90
88
86
84
82
80
S= 0
S=0.3
S=0.7
S=1.0
70
90
110
130
110
170
190
КАНАЛЫ

Рис.2.

максимального правдоподобия, который при $N_i \gg 1$ равносилен определению α из условия минимизации функционала

$$\varphi(N_i;\alpha) = \sum_{i=1}^{N} \frac{1}{N_i} \cdot \left(N_i - MN_i(\alpha) \right)^2 . \qquad (14)$$

При численном решении задачи (14) применялся метод линеаризации [7,8].

Для соединения $Cd\,Fe_{1/2}Nb_{1/2}O_3$ лучшее приближение экспериментального спектра получено с моделью решетки, в которой ионы одинакового сорта образуют цепочки, располагающиеся по плоскостям (110). В структуре имеется два типа неэквивалентных узлов ($r = 2$).

С первой посылкой об изомерном сдвиге решение обратной задачи дает оценки параметров $\widetilde{N}_\infty = 99900$, $\widetilde{A} = 34500$, $\widetilde{\Delta E} = 12{,}6$, $\widetilde{X} = 131{,}6$, $\widetilde{\Gamma} = 16{,}3$ с оценками стандартов $\sigma\widetilde{N}_\infty = 30$, $\sigma\widetilde{A} = 200$, $\sigma\widetilde{\Delta E} = 0{,}1$, $\sigma\widetilde{X} = 0{,}1$, $\sigma\widetilde{\Gamma} = 0{,}4$. Значение функционала (14) при указанных оценках и $S = 0{,}4$ равно 375 при $N = 256$.

Для соединения $Sr\,Fe_{2/3}W_{1/3}O_3$ с описанной структурой решетки и первой версии об изомерном сдвиге оценки параметров $\widetilde{N}_\infty = 99600$, $\widetilde{A} = 18200$, $\widetilde{\Delta E} = 10{,}7$, $\widetilde{X} = 127{,}7$, $\widetilde{\Gamma} = 20{,}5$, $S = 0{,}9$ с тем же уровнем ошибок. Минимальное значение функционала равно 311. Полученные приближения экспериментальных спектров приведены на рис. 1 (кривые 2).

В таблице П приведены значения вероятностей $P(n)_K$ исследованных соединений для лучшего приближения экспериментальных спектров. Для сравнения в последних двух столбцах даны вероятности $P(n)_K$ для полностью неупорядоченного распределения.

Перейдем к обсуждению второго аспекта интерпретации экспериментальных мессбауэровских спектров. Задачу выбора первой или третьей версии об изомерном сдвиге можно свести к задаче проверки статистических гипотез. Действительно, выдвигая нулевую гипотезу $H_0: \Delta = 0$, что соответствует первой посылке об изомерном сдвиге, мы имеем двухстороннюю альтернативу $H_1: \Delta \neq 0$, что соответствует третьей посылке. При проверке гипотезы H_0 против H_1, мы можем воспользоваться, например, критерием основанным на распределении

Таблица II

$P(n)_K$	$CdFe_{1/2}Nb_{1/2}O_3$	$SrFe_{2/3}W_{1/3}O_3$	$CaFe_{2/3}W_{1/3}O_3$	Неупорядоч. распределен. $AB'_{2/3}B''_{1/3}O_3$	Неупорядоч. распределен. $AB'_{1/2}B''_{1/2}O_3$
$P(0)_0$	0.0I6	0	0.00I	0.00I	0.0I6
$P(I)_1$	0.I09	0	0.0II	0.0I6	0.094
$P(2)_1$	0.I68	0	0.0II	0.066	0.I87
$P(2)_2$	0.I0I	0.234	0.I64	0.0I6	0.047
$P(3)_0$	0.074	0	0.003	0.088	0.I25
$P(3)_{\sqrt{3}}$	0.246	0.II7	0.I80	0.I32	0.I88
$P(4)_1$	0.I39	0.2I5	0.052	0.263	0.I88
$P(4)_2$	0.068	0.468	0.363	0.066	0.047
$P(5)_1$	0.070	0.I54	0.I92	0.263	0.094
$P(6)_0$	0.009	0.0I2	0.023	0.088	0.0I6

статистики

$$t = \frac{|\tilde{\Delta}|}{\sigma\tilde{\Delta} \cdot \sqrt{\frac{\Phi(N_i;\tilde{\alpha})}{N-7}}}, \qquad (15)$$

имеющей асимптотическое распределение Стьюдента. Сам критерий при этом имеет вид

$$t > t_\varepsilon, \qquad (16)$$

где ε – выбранный уровень значимости, t_ε – ε – процентная точка t – распределения, т.е. гипотеза H_1 отвергается в пользу H_0 при выполнении неравенства (16) и не отвергается в противном случае.

Для рассмотренных задач гипотеза H_0 не отвергается практически на любом разумном уровне значимости. Иными словами, результаты экспериментов не противоречат справедливости посылки о том, что изомерный сдвиг не зависит от окружения мессбауэровского ядра.

Изложенная выше методика обработки экспериментальных мессбауэровских спектров позволяет корректно оценить важный физический параметр – температуру магнитного упорядочения. Эта температура в окислах со структурой перовскита, обычно, рассчитывается по схеме Гуденафа-Гильо [9,10]. В основе этой схемы лежит предположение о том, что при замещении частиц магнитных ионов немагнитными первый ион участвует в магнитном упорядочении, если он имеет не менее двух магнитных соседей. Следоватсльно, энергия обменного взаимодействия, а значит и температура магнитного упорядочения, пропорциональны количеству магнитных взаимодействий или количеству взаимодействующих пар $Fe-O-Fe$ в исследуемых перовскитных соединениях. Однако, в работах, где рассчитывались температуры магнитного упорядочения сложных перовскитных соединений [2,11] не получены удовлетворительные согласия с ее экспериментальными значениями. По-видимому, дело в том, что при теоретическом расчете температуры магнитного упорядочения используются значения вероятностей распределения ионов B' и B'' по двум магнитным подрешеткам, вычисленные по схеме Бернулли, Кроме того, не учитывается влияние на температуру сверхобменного взаимодействия типа $B'-O^{2-}-B''-O^{2-}-B'$.

Ниже предлагается новый подход к оценке температуры маг-

нитного упорядочения перовскитных соединений сложного состава. Число эффективных магнитных связей на ион Fe, участвующего в магнитном упорядочении, в отличие от работ [2, II], будем определять по формуле

$$n_{эфф} = \sum_{n=2}^{6} n \sum_{k=0,1,\sqrt{3},2} P(n)_k, \qquad (I7)$$

где $P(n)_k$ – вероятности, найденные из экспериментальных мессбауэровских квадрупольных спектров по оценке параметра S. Используя $P(n)_k$, приведенные в таблице II, вычислим значения $n_{эфф}$ для исследуемых соединений

$n_{эфф}(Sr) = 4.39$, $n_{эфф}(Ca) = 3.66$

Для сложного перовскитного соединения $PbFe_{2/3}W_{1/3}O_3$, в котором распределение ионов Fe и W полностью неупорядоченное $n_{эфф}(Pb) = 3.98$. Сравним температуры магнитного упорядочения соединений [I2] и числа эффективных магнитных связей на ион Fe относительно соответствующих характеристик $PbFe_{2/3}W_{1/3}O_3$ [II]:

$$\frac{T(Sr)}{T(Pb)} = \frac{393^\circ K}{363^\circ K} = 1.08, \quad \frac{n_{эфф}(Sr)}{n_{эфф}(Pb)} = \frac{4.39}{3.98} = 1.10,$$

$$\frac{T(Ca)}{T(Pb)} = \frac{329^\circ K}{363^\circ K} = 0.89 \quad \frac{n_{эфф}(Ca)}{n_{эфф}(Pb)} = \frac{3.66}{3.98} = 0.90$$

Итак, изложенный подход показывает, что температура магнитного упорядочения сложных перовскитных соединений пропорциональна $n_{эфф}$ и может быть рассчитана по формуле

$$T = \frac{n_{эфф}}{n_{эфф\ неуп}} T_{неуп}, \qquad (I8)$$

где $n_{эфф\ неуп}$ и $T_{неуп}$ – число эффективных магнитных связей и температура магнитного упорядочения соединения, в котором ионы B' и B'' полностью неупорядочены. Приведенные числовые данные иллюстрируют близость теоретических и экспериментальных значений температуры магнитного упорядочения.

ЛИТЕРАТУРА

I. Bell R.O. J.phys. Chem. Solids, 29, 69, I968.

2. Uchino K., Nomura S. Ferroelectrics, I7, 505, I978.

3. Галкин В.Я., Горьков В.П., Кузьмин Р.Н., Ройг Нуньес Х.Х., Шагдаров В.Б. Сб."Некоторые вопросы автоматизированной обработки и интерпретации физических экспериментов",вып.2, Изд-во МГУ,I973.

4. Кузьмин Р.Н.,Лоссиевская С.Н., ФММ,29,569,I970.

5. Gorskii V.S. Zs. Sowjet Union, 8, 443, I935.

6. Bragg W.L., Williams E.F. Proc. Roy. Soc., A. I52,29I,I935.

7. Соколов Н.С., Силин И.Н. Препринт ОИЯИ,Д-8I0,Дубна,I96I.

8. Библиотека программ на ФОРТРАНе Д-520,т.I,Дубна,I970.

9. Goodenough J.B., Wickham D.G. Groft J.W. J. Phys. Chem. Solids, 5, I07, I958.

I0. Gilles M.A. J.Phys. Chem. Solids, I3, 33, I960.

II. Смоленский Г.А., Исупов В.А.,Крайник Н.И., Аграновская А.И. Из. АН СССР,сер.физ.,25, I333,I96I.

I2. Иванова В.В.,Капышев А.Г.,Веневцев Ю.Н. Из. АН СССР Неорганические материалы,6, I68, I970.

ИНТЕРАКТИВНОЕ РЕДАКТИРОВАНИЕ ПРОГРАММ И ПОДГОТОВКА ЗАДАЧ НА ЭВМ ЕС-1040

А. Арато, И. Шаркади-Надь, Ф. Телбис
Центральный институт физических исследований, Будапешт

АННОТАЦИЯ

В Центральном институте физических исследований ВАН разработана терминальная сеть для улучшения эффективности разработки программного обеспечения. В центре сети находится ЭВМ ЕС-1040, к ней подключена мини-машина ТРА-70, как буферный процессор. Интерактивными терминалами являются дисплеи типа VT-340 и матричные печатающие устройства типа DZM-180. Один дисплей с матричным печатающим устройством служит главным консолем машины ЕС-1040.

Система осуществляет три функции:
а. Интерактивное редактирование текста.
б. Обеспечивает возможность посылки задания для пакетной обработки.
в. Вывод результатов заданий терминал.

В докладе рассматриваются опыты эксплуатции интерактивной системы CEDRUS и некоторые методы повышения надежности системы.

ABSTRACT

In order to improve the efficiency of software production a terminal system was developed in the Central Research Institute for Physics. There is an ES-1040 computer in the centre of the system with a TPA-70 mini machine as front-end-processor. VT-340 type displays are used for interactive terminals and DZM-180 type matrix printers are used for hard copy devices. One of the displays together with a matrix printer is used as the master console of the ES-1040 computer.

The system has three services:
a./ Interactive text editing
b./ Job submission to the batch processing
c./ Fetching the results of jobs

Some reliability questions of the system are discussed in the paper.

Введение

В центральном институте физических исследований была сдана в эксплуатацию первая очередь локальной сети ЭВМ CEDRUS. Это сокращенное название системы "Conversational EDitor and Remote Users' Support". В центре сети стоит вычислительная машина ЕС-1040 крупной конфигурации /1 Мбайт оперативной памяти, 24 устройства накопителей на магнитных дисках емкостью 7,25 Мбайт, 8 устройств накопителей на магнитных лентах, 3 карточных ввода, 3 АЦПУ, 1 перфоратор/. В первой очереди системы CEDRUS был осуществлен редактор текста. В настоящее время в системе могут работать одновременно десять пользователей у дисплейных установок. Дисплеи подключены к буферному процессору, осуществленному мини-машиной типа ТРА-70. Терминалы находятся или вблизи машины, или на расстоянии нескольких сотен метров.

Во второй очереди осуществления системы CEDRUS будут подключены удаленные мини-машины типа ТРА/i, ТРА 1140, ТРА-70 и микромашины, построенные на основе микропроцессоров Intel 8080 и т.д. Эти мини и микромашины, которые служат для измерения в экспериментах, смогут передавать через сеть файлы данных для дальнейшей обработки в центральной вычислительной машине.

В докладе будут рассмотрены некоторые вопросы осуществления интерактивного редактора и опыта полугодовой эксплуатации системы.

Способ связи машины

Для стыковки общей шины машины ТРА с каналом машины ЕС был разработан адаптер. Этот адаптер дает возможность эмуляции 128-и физических адресов на мультиплексном или селекторном канале. Эмулированные физические устройства могут принадлежать к одной из двух типовых групп. Первая группа, так называемая shared - вторая - multiple shared. Внутри одной группы с помощью эмулирующей программы могут быть эмулированы физические устройства с разным составом команд. Адаптер содержит встроенные приспособления для on-line тестирования. Максимальная скорость передачи данных из одной ЭВМ в другую - 250 000 байт в секунду.

В системе CEDRUS - оператор подключен к селекторному каналу. Программа эмулирует 16 физических адресов. Из них 8 перфокарточных вводов ЕС-1062 и 8 АЦПУ ЕС-7033. Из этих устройств одна пара ввод-вывод служит для обмена информацией между программами редактора текста, работающими в центральной машине и в буферном процессоре. Две пары устройств ввод-вывод сгенерированы в операционной системе OSMVT как составные консоли мультиконсольной системы. Один составной консоль осуществлен дисплеем и матричным печатающим устройством и заменяет медленнодействующий механический консоль ЭВМ ЕС-1040. Второй составной консоль служит для опроса состояния задач, посланных от терминалов. Следующая пара устройств дает возможность использовать перфоленточные механизмы машины ТРА из программ ЕС ЭВМ.

Услуги системы

Пользователи, которые имеют учетные номера в системе и имеют область на on-line дисках, могут редактировать свои программы, данные или любую текстовую информацию. После редактирования они могут посылать задания для пакетной обработки или печатать текст на АЦПУ с обработкой формата вывода.

Ниже коротко перечислены команды:

- Команды под файлами:
  - USE — перепись постоянного файла в рабочий файл
  - JOIN — присоединение постоянного фалйа к рабочему
  - SAVE — перепись рабочего файла в постоянный
  - SCRATCH — стирание постоянного файла
  - CLEAR — стирание рабочего файла
  - PRINT — печатание рабочего файла
  - PUBLISH — то же самое с выравниванием полей и управлением формата
- Команды редактирования в рабочем файле
  - ADD — добавление строк
  - DELETE — стирание строк
  - REPLACE — замещение строк
  - CORY — дублирование строк
  - MOVE — перемещение строк
  - LIST — листинг строк на дисплее
  - CHANGE — замещение стринга в одной или нескольких строках
  - SEARCH — поиск строк, содержащих определенный стринг
  - EDIT — редактирование внутри строки
  - RENUMBER — перенумерация строк
- Команды пакетной обработки
  - SUBIT — пересылка задания для пакетной обработки
  - FETCH — получение результатов на терминале
  - STATUS — запрос о состоянии заданий
  - PURGE — стирание результатов с диска
- Команды администрации
  - LOGIN — вступление в систему
  - LOGOUT — уход из системы
  - HELP — запрос системы о помощи, как ее использовать
  - TALK — передача сообщения оператору или терминалам

Надежность системы

Редактор текста CEDRUS работает на машинах ограниченной надежности, поэтому большое внимание было уделено тому, чтобы как можно меньше работы терялось у пользователей. Для этой цели все таблицы и данные пользователей записываются на дисках как только в них произошло изменение. С помощью такой стратегии при выходе из строя операционной системы "зависания" теряется только последняя команда у одного пользователя или и того меньше. На практике таких ситуаций бывает от одной до трех за 12 часов работы.

Другой мерой для увеличения надежности являются программы избавления от ошибок диска типа ЕС-5052. Постоянные файлы записываются канальной программой, которая содержит проверку только что записанной информации. При ошибке выдается сообщение на нужный терминал. С помощью такой канальной программы ошибки в постоянных файлах сведены к минимуму.

Специальной программой избавления от ошибок диска освобождаемся от сбоев в рабочем файле. Рабочий файл организован как файл непосредственного доступа /DA/. Очень часто встречаются ошибки типа "запись не найдена" /NO Record Found NRF/. При такой ошибке специальная программа читает запись, которая на самом деле не потерялась, только идентификатор не читается. Эта программа после считывания записи обновляет идентификатор и работа автоматически продолжается. Такие сбои из-за очень активного использования рабочего файла встречаются раз в два-три часа работы.

Итоги:

Как показывает опыт эксплуатации системы CEDRUS, из-за неиспользования перфокарт и из-за удобных интерактивных команд эффектинвость разработки программ увеличивается в 2-3 раза. Пользователи легко выучивают команды и при синтаксических и семантических ошибках получают понятные сообщения о своих ошибках.

Поскольку команды имеют свободный формат и могут быть сокращены, система удовлетворяет требованиям и новичков и пользователей с большой практикой работы в системе.

Литература:

1 R.D.Russel-P.Sparman-M.Krieger: ORION - The OMEGA remote interactive on-line system.
Proc. International Computing Symposium 1973.p.143
/North Holland, 1974/

2 A.Arató-I.Sarkadi Nagy-F.Telbisz: A local network of software development.
Proc. of COMNET'77 Budapest 1977. Vol. 1. p. 227.

ИЗМЕРЕНИЕ И МОДЕЛИРОВАНИЕ ИНТЕРАКТИВНОЙ ТЕРМИНАЛЬНОЙ СИСТЕМЫ CEDRUS

А. Арато, И. Шаркади-Надь, Ф. Теблис

Центральный институт физических исследований, Будапешт

АННОТАЦИЯ

В работе описывается метод измерения времени ответа интерактивной программы, работающей в машине ЕС, с помощью программ и схем буферного процессора. По результатам измерения была создана модель программы на языке GPSS. Правильность работы модели доказывалась сопоставлением модели с результатами измерения. Даны некоторые результаты использования проведенной модели. Результаты измерения и моделирования приведены в виде графиков.

ABSTRACT

A method is discussed for measuring response time of an interactive program running in ES machine with the software and hardware of the front end processor. A simulation model of the program was written in GPSS on the basis of some measuring results. The verification of the model was made by comparison of the measuring results. Some applications of the verified model is discussed. The measurement and simulation are illustrated with histograms.

1. Введение

В интерактивной системе, какой является и редактор текста CEDRUS |5,6| одной из важнейших характеристик является время ответа. Временем ответа интерактивной системы называется то время, которое проходит от момента посылки команды или запроса, до того момента, когда появляется первый символ на экране, или бумаге терминала. Другими словами время ответа характерузет быстроту реакции системы. Конкретное определение времени ответа в редакторе текста CEDRUS будет дана позже. Время ответа определяется скоростью каналов, эффективностью программ, скоростью машин, оперативной памяти, вспомогательной памяти и т.д.

Величины времени ответа могут быть очень разные в различных системах |1|. Интерактивные системы отличаются более строгими допустимыми максимальными временами реакций системы. В диалоговых системах ограничивающим фактором является человеческое терпение. Человек выдерживает не более 5-и секунд задержки ответа на его запрос. Это максимальное значение относится к тем запросам, ответы на которые должны прийти быстро. Есть такие команды, выполнение которых тянется более долго. Времена ответа на эти команды могут доходить до нескольких минут /например перепись нескольких тысяч записей из одного файла в другой/. Предел интерактивности, на этом и кончается. Получение результата, в пакетной обработке заданий уже часто выходит из этого предела времени ответа.

Нужно определить еще одно время, которое резко увеличивает время ответа. Это время восстановления системы из опшибок. Примером такой ошибки может служить ошибка данных на дисках или ошибка при передаче данных через линии связи, если возможно их исправить повторением. Анализ таких задержек тоже будет проделан по отношению интерактивного редактора CEDRUS.

2. CEDRUS как двухмашинный комплекс

Редактор текста CEDRUS осуществлен на двух машинах |7|. Одна часть программного обеспечения работает в машине ЕС а другая часть работает в буферном процессоре /в мини машине ТРА/. Задачи между машинами распределены по такому принципу, чтобы каждая машина делала ту часть работы, которая более

подходит к конструкции данной ЭВМ. Машина ЕС имеет много вспомогательной памяти, очень развитую операционную систему и методы доступа к фамилям, но имеет очень бедную систему прерываний. Эта машина очень хорошо подходит для последовательной работы и предназначена, в первую очередь для пакетной обработки. Минимашина может иметь не так много вспомогательной памяти, но имеет развитую, многовекторную систему прерывания многих уровней. Мини машина хорошо подходит для работы в реальном масштабе времени.

В системе CEDRUS было выбрано такое распределение задач между машинами: Все манипуляции над файлами на дисках осуществляет машина ЕС. Редактирование на уровне строк происходит тоже в большой ЭВМ. Редактирование внутри строк делает минимашина. Ввод и вывод, связанный с терминалами - это задача минимашины. Синтаксический анализ команд напечатанных пользователями происходит тоже в минимашине. Большая машина получает от буферного процесса уже предварительно обработанные буфера. Большая машинна выполняет команды, указанные в этих буферах связи строго последовательно. Очередь буферов связи образовывается в машине ТРА. Структуру двухмашинного комплекса иллюстрирует следующий рисунок /Рис. 1/.

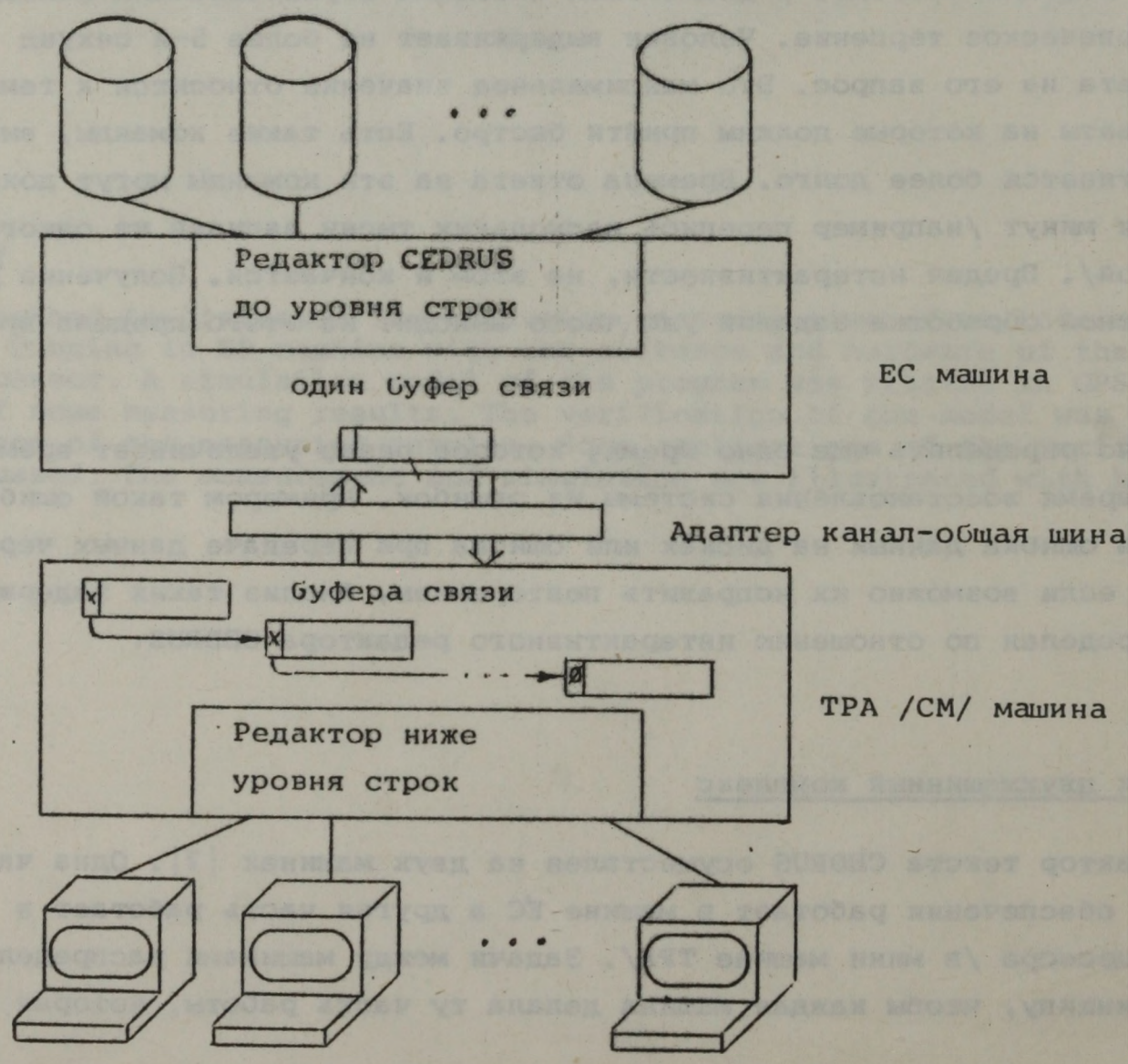


Рис. 1

Структура двухмашинного комплекса

После того, как известно строение системы, можно приступить к определению времени ответа в ней.

3. Время ответа в редакторе CEDRUS

Полное время ответа по определению в введении состоит из следующих составляющих:

$$T=(t_{CA}+n.t_{CA}+m.t_{ред}+2t_{кан}+T_{очер.}+T_{EC}) \ .b$$

где T полное время ответа

t_{CA} время синтаксического анализа одной команды

$t_{ред}$ время редактирования в строке в мини машине

$t_{кан}$ время прохождения буфера связи через канал

n,m стохастические коэффициенты равномерного распределения от 0 до максимального числа терминалов. Сумма $n+m \leqslant$ максимального числа терминалов. /зависит от типа команд/

b число обмена буферами связис

$T_{очер}$ Это задержка в очердеи буферов связи

T_{EC} время выполнения команды в машине ЕС

Время синтасического анализа составляет не более 1 мсек. Также обстоит дело и с временем редактирования. Эти величины времени ответа определяются главным образом скоростью процессора мини машины. Скорость передачи данных из мини машины в ЕС ЭВМ составляет 250 000 байт в секунду. В редакторе используются буфера связи "длиной" в 512 байт. Передача буфера "вверх" и "вниз" происходит за 4 мсек. Времена $T_{очер}$ и T_{EC} не могут быть просто выражены аналитически. Они содержат много стохастических переменных сложных распределений. Главной задержкой в этих составляющих является\время доступа к записям на дисках. Минимальное время доступа к одной записи на диске типа ЕС-5050 является 10 мсек. Таких доступов за время T_{EC} нужно делать несколько. Из предыдущего видно, что можно пренебрегать некоторыми составляющими времени ответа. Если пренебрегать временем, которое нужно на выполнение команд реадктора арифметическим устройством машины ЕС в составляющих $T_{очер}$ и T_{EC}, то время ответа можно считать равным:

$$T = (T_{очер}+T_{EC}).b$$

При измерении системы составляющие времени ответа t_{CA} и $t_{ред}$ не учитываются из-за способа измерения. Коэффициент b считается равным одному. Это тоже вытекает из способа измерения. При моделировании редактора используется формула

$$T = T^{M}_{очер} + T^{M}_{EC}$$

если сравнить ее с формулой, использованной для описания измерения:

$$T = 2t_{кан} + T^{и}_{очер} + T^{и}_{ЕС}$$

то видно, что они различаются на 4 мсек и на разницу $T^{и}_{очер} - T^{м}_{очер}$ и $T^{и}_{ЕС} - T^{м}_{ЕС}$, т.е. на время, которое нужно для передачи данных и счета в процессоре ЕС ЭВМ.

4. Время восстановления системы из сбоев

Когда в системе присутствуют ненадежные схемные и механические элементы, тогда возникают сбои в статистические моменты. Кроме сбоев схем можно говорить о собоях операционной системы /мертвые точки/. О методах избежания мертвых точек операционной системы будет написано в другой работе. Из сбоев операционной системы /"зависания"/ машины ЕС можно восстановить систему только с новой загрузкой программ в ЭВМ. Это время настолько велико, что ситуацию нужно рассматривать как аварийную, выходящую из пределов интерактивности.

Другие типы сбоев, ошибки на дисках могут быть рассмотрены с точки зрения времен ответа. Были измерены статистические параметры одного вида ошибок на дисках. Механизм ошибки состоит в том, что при многократной записи данных методом непосредственно доступа, идентификатор записи повреждается. В среднем после каждой 18000-ой записи происходит сбой. В редакторе CEDRUS по отношению этих сбоев, наибольшей опасности подвергаются так называемые записи безопансоти. Сюда записываются все таблицы системы после каждой выполненной команды, если в этих таблицах произошло изменение. Таким образом достигается, то что при "зависании"теряется только последняя команда у одного пользователя или и того меньше. Восстановление работы редактора из этого сбоя происходит обновлением идентификаторной части записи. Время восстановления состоит из следующих составляющих частей

$$T_{В} = 16(t_{рек} + 16t_{об}) + t_{опер} + (2t_{open} + 2t_{close} + 5t_{об})$$

где
- $T_{В}$ время восстановления
- $t_{рек}$ время выполнения команды диска рекалибрации
- $t_{об}$ время одного оборота диска
- t_{open} время открывания файла
- t_{close} время закрывания файла
- $t_{опер}$ время реакции оператора на сообщение системы

Первый член формулы описывает попытки операционной системы избавиться от ошибки. Другой член в скобках описывает время, необходимое для избавления от ошибки специальной программы. Время реакции оператора и сообщения системы могут быть исключены, если операционная система OS эксплуатируется без возможности динамической реконфигурации устройств /SWAP OFF/. В этом случае T_B будет составлять порядка 30 секунд. Если организовать программы приложения, то это время может быть сведено к нескольким секунадм, которое не превышает допустимый предел интерактивности.

5. Программные и схемные методы измерения

Измерение системы проводится для определения входных данных моделирования с одной стороны, а с другой стороны для проверки правильной работы модели.

Измерение статистических параметров сбоев диска было определено посредством тестовой программы. Эта программа создала идентичные условия подобные тому, как редактор CEDRUS записывает записи безопасности.

Число доступов к диску во время выполнения таких системных функций как открытие и заркытие файлов, выделение или освобождение области диска фалйов и т.д. могут быть получены программой операционной системы /General Trace Facility/. Функция распределения времен доступа к записи на диске может быть измерена качественно отдельной программой.

Остальные входные распределения и времена ответа были измерены комплексом схем и программ буферного прцоессора. Адаптер буферного процессора дает возможность эмулировать независимые физические одреса на селекторном канале машины ЕС. Один из этих адресов служил для целей измерения. Времена измерялись кварцовым таймером и специальной программой прерывания буферного процессора. Посокльку измерение и моделирование ограничивается только рассмотрением программы, работающей в машине ЕС, как это было указано в разделе 3., то метод измерения извне, схемами и программами буферного процессора дает/идеально точные результаты. Условия измерения были неидеальны только из-за того, что у буферного процессора не было вспомогательной памяти. Данные измерения записывались на магнитную ленту в машине ЕС. Это было программа обычной утилити, потому что данные из буферного процессора передавались через эмулированный считыватель перфокарт. Для уменьшения влияния на результаты измерения, эта программа утилити имела меньший приоритет, чем программа редактор CEDRUS, и данные были собраны в большие блоки. Структуру измерения описывает следующий рисунок. /Рис. 2/.

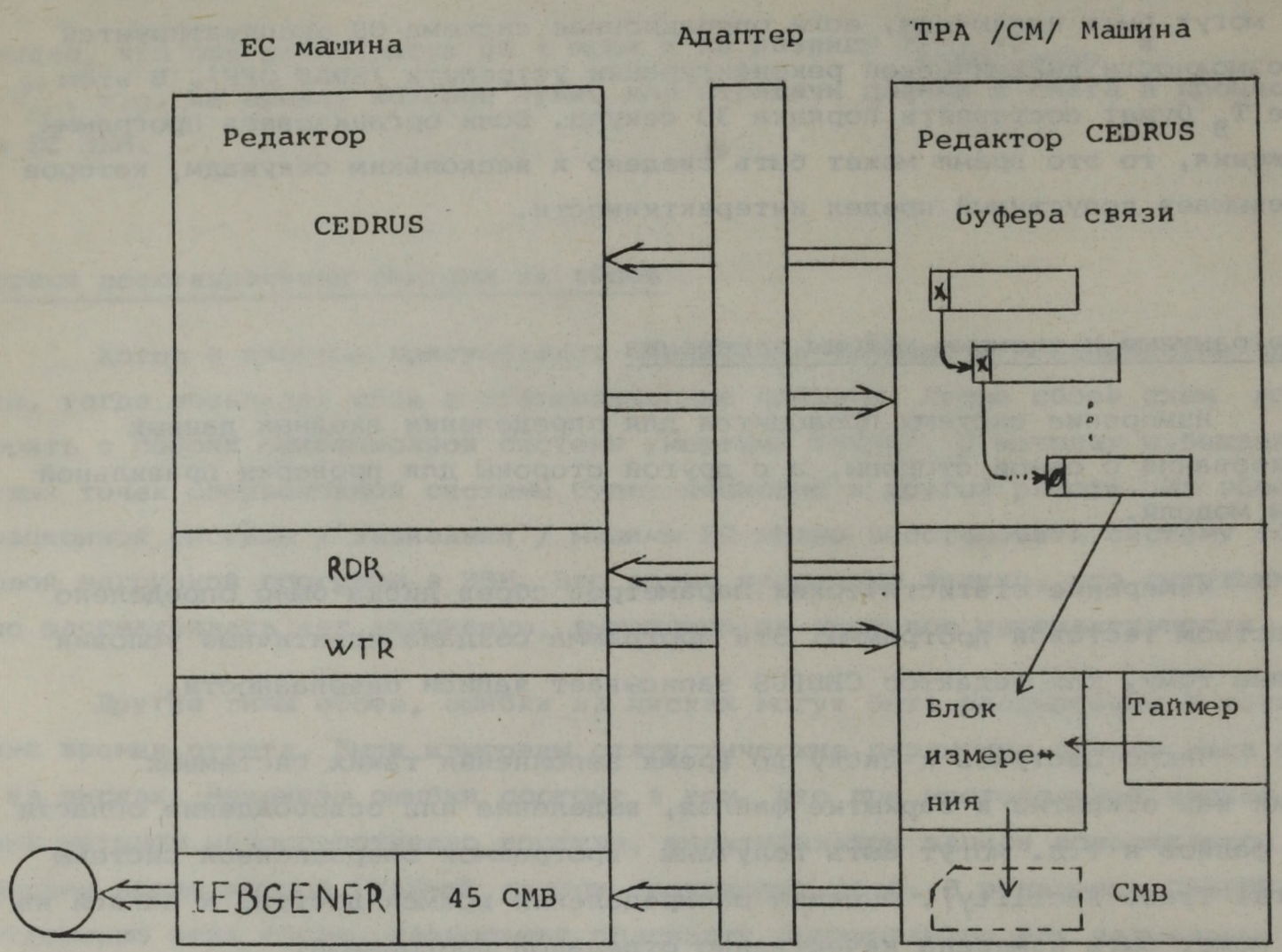


Рис. 2

Структура измерения программы ЕС

Блок измерения регистрирует время прихода в очередь каждого буфера связи, и время, которое происходит от этого начального момента до момента возвращения буфера связи из ЕС ЭВМ. Времена измеряются кварцовым таймером буферного процессора. Из-за некоторых пренебрежений при моделировании, которые описаны в разделе 3., цикл таймера установлен 0,1 секунды. После прихода буфера связи из ЕС ЭВМ.блок измерения составляет очередную запись измерения /CMR-CEDRUS Measurement Record/. Пять таких записей составляет одну мнимую карту, блок измерения /CMB-CEDRUS Measurement Block/. Эти карты блокирует утилити /IEBGENER/ в блоки с размером 3600 байт. Для того, чтобы из-за задержек записи информации на магнитную ленту не терялись данные измерения, в буферном процессоре отделено место в памяти для шести мнимых карт /CMB/. При максимальной скорости обмена буферов связи между машинами /25 за секунду/ записи на магнитную ленту производится малой частотой /1 запись за 9 секунд/. Если учитывать, что магнитные ленты подключены к другому каналу, чем диски, то видно, что этот метод сбора данных измерения вносит очень маленькую погрешность. Структуру данных измерения иллюстрирует следующий рисунок: /Рис. 3/

Запись измерения CMR:

c	n	t	s	o_1	o_2	r	stime	interval

код	название	номер байта	объяснение
c	COMMAND	Ø	код команды в буфере связи
n	USERNO	1	номер пользователя в системе
t	TERMINALNO	2	номер терминала
s	ERRORCODE	3	код ошибки
o_1	OPTION1	4	добавочный код команды 1
o_2	OPTION2	5	добавочный код коменды 2
r	RECORDS	6	число записей затронутых командой
stime	STARTIME	8	время прихода буфера в очередь
interval	INTERVAL	11	время возвращения буфера из ЕС

Блок измерения CMB:

CMR	CMR	CMR	CMR	CMR	№	CARD NO
14б	14б	14б	14б	14б	2б	8б

Рис. 3

Структура данных измерения

Данные измерения, записанные на магнитную ленту обрабатываются после измреения с помощью off-line программой. Выходной формат этой программы совпадает со входным форматом моделирующей программы. Входные распределения расчитываются из первых восьми полей записей измерения. Последнее поле INTERVAL служит для проверки правильной работы модели.

6. Моделирование редактора CEDRUS

Целью моделирования программы редактора работающей в машине ЕС, является опердеелние времен ответа, если бы в системе работали более быстрые диски. Кроме того получить ответ на то, сколько терминалов может работать одновременно при этих быстрых дисках и при терпимых временах ответа. В последнюю очередь моделированием можно получить ответ на то, какие качественные и количественнные изменения нужно делать в оригинальной программе, для уменьшения времен ответа.

Для моделирования был выбран язык GPSS |3|. Преимущество этого языка по отношению к языкам общего назначения например к ФОРТРАН-у или даже по отношению более специализированного языка SIMULA 67 состоит в его компактности, легкости статистических расчетов и удобном встроенном редакторе вывода обычной и графической информации. 5000 строк программы оригинала моделироваолось 500-ми строками GSSP. Для нестатистических расчетов в языке GPSS имеется блок HELP, через который в модель можно писать любые подпрограммы на любых языках. Сравнительный анализ языков моделирования более подробно содержится в книге |2|.

Принцип построения модели состоялся в том, чтобы все задеркжки времени сводились к задержке из-за доступа к диску. Модель построена структурно, широко используются возможности вызова подпрограммы, и символьного программирования. Введено структурная форма типа CASE OF или похоже как на ФОРТРАН-е ASSIGN TO. В модели содержится единственный блок ADVANCE следующей формы:

ADVANCE FN$ACCES /ACCESstime/

где функция FN$ACCES является входным распределением модели, описывающим вероятность доступа к одной записи на диске. В этой фнукции учитывается не только скорость вращения диска, сокрость движения позиционера, но и задержка в очереди к каналу в супервызоре машины ЕС. Изменением этой функции можно моделировать влияние на время ответа с одной стороны от скорости диска, сдругой стороны влияние от увеличения общей загруженности ЕС ЭВМ другими программами. Буфера связи генерируются фиункцией FN$LINKB /LINK Buffer/, таблица которой получается из программ обработки данных измерения. Распределение вероятности размеров файлов описывает функция FN$DSSIZ /Data Set SIZe/. Распределения типа команд и номеров терминалов описывают функции FN$COMNO /COMmand NO/ и FN$TERNO /TERminal NO/ соответственно. Для моделирования команды SHOW ACCOUNT, понадобилась функция FN$VTOCS /VTOC Size/. Сообщения на терминал при синтаксических или логических ошибках выдаются очень подробно. Также много информации можно получить при использовании команды HELP. Эти данные хранятся на диске машины ЕС. Поиск нужной строки в этих фалйах происходит последовательно, поэтому для моделирования этих команд используются функции FN$ESEAR /Error SEARch/ и FN$HSEAR /Help SEARch/. В редакторе CERUS лимитровано число карт в задании /JOB/ в команде SUBMIT. Это учитывает функция FN$SCOUN /Submit COUNt/. Оверлейную структуру программы описывает таблица FN$OVLY /OVerLaY/.

Выходной информацией модели явлется серия таблиц и графиков распределения времени ответа для каждого типа команд в отдельности и обобщенно.

7. Отладка модели

Наиболее трудной задачей моделирования является отладка модели. После простой синтаксической отладки стоит вопрос правильно-ли работает модель. сли возможно подробное измерение модели в разных краевых ситуациях, тогда можно придерживаться следующего метода.

Нужно измерить поведение системы в краевых условиях. На примере редактора текста CEDRUS это означает, что вступает в систему один единственный пользователь и выдает команду например SHOW ACCOUNT. т.е. получить листинг постоянных файлов на данном учетном номере. Разница во времени ответа между первой и последующей командой описывает время нужное для смены оверленйного слоя.

В модели можно создать аналогичные условия, если функцию FN$COMNO описывающую распределение команд определить таким образом:

```
COMNO   FUNCTION    RN2,D2
.ØØØ,1/.999,21
```

где 21 это номер команды SHOW ACCOUNT. При этом буфера связи с командой 21 нужно генерировать так редко, чтобы они не должны были ожидать друг друга в очереди. Для примера: генерировать буфера связи через фиксированные интервалы времени. Зная, что одно деление таймера модели было выбрано равным 1/1ØØ секунды, блок GENERATE на языке GPSS будет иметь вид:

```
GENERATE 2ØØØ,Ø,,,,1ØH
```

Результаы отладки приведены на следующих рисунках: Пример протокола отладки команды приведен на рис. 4. На этом рисунке видно, что на терминале была выдана команда SHOW ACCOUNT 16 раз, для различных учетных номеров. Между апострофами задается название файла, который не будет найден и поэтому на одной странцие протокола помещается все 16 команд. В том случае, когда не дан четырехсимвольный стринг используется тот учетный номер, под которым вступил пользователь данного терминала. Выбраны такие учетные номера, которые имеют файли на различных пакетах дисков. На рисунке 5. приведен листинг результата измерения. Одна строка соответствует одной записи измерения /CMR/. Выделена та часть листинга, которая соовтетствует протоколу на рис. 4. Номер команды SHOW ACCOUNT = 21, а команды TIME = 5. Графа RECORD /число сорок/ носит информацию только при командах PRINT = 7, PUBLISH = 8, USE JOIN = 15, SAVE = 17 и SCRATCHSAVE=19. В графе RESPONSETIME /вермя ответа/ одна единица соответствует 1/1Ø секунды. На рис. 6 показан график результата моделирования при условии, что команда SHOW ACCOUNT выдается только с одного терминала и что у других терминалах не работают. На оси абцисс цифры соответствуют секундам времени ответа. На оси ординат указана частота появления времен ответа, в данном интервале времени /в процентах/. Более высокие времена ответа при моделировании

```
  COMMAND? SHOW ACCOUNT 'THESE FILES...';
FREE DISK SPACE:     7 QUANTA

  COMMAND? SHOW ACCOUNT '...WILL NOT BE FOUND';
FREE DISK SPACE:     7 QUANTA

  COMMAND? SHOW ACCOUNT 'ONLY FOR TEST';
FREE DISK SPACE:     7 QUANTA

  COMMAND? SHOW ACCOUNT 'TEST';
FREE DISK SPACE:     7 QUANTA

  COMMAND? SH A 'TEST';
FREE DISK SPACE:     7 QUANTA

  COMMAND? SH A KAFF 'TEST.ONLY';
FREE DISK SPACE:    65 QUANTA

  COMMAND? SH A KMDD 'TEST.ONLY';
FREE DISK SPACE:   180 QUANTA

  COMMAND? SH A KMAB 'FOR TEST';
FREE DISK SPACE:   138 QUANTA

  COMMAND? SH A KMCI 'FOR TEST';
FREE DISK SPACE:    53 QUANTA

  COMMAND? SH A KRMO 'TEST ';
FREE DISK SPACE:    38 QUANTA

  COMMAND? SH A KMD1 'THIS IS USER11';
FREE DISK SPACE:     1 QUANTA

  COMMAND? SH A KMEL 'THIS IS USER12';
FREE DISK SPACE:   124 QUANTA

  COMMAND? SH A KREC 'THIS IS USER13';
FREE DISK SPACE:    25 QUANTA

  COMMAND? SH A KATA 'THIS IS USER14';
FREE DISK SPACE:   150 QUANTA

  COMMAND? SH A KSCV 'THIS IS USER15';
FREE DISK SPACE:   150 QUANTA

  COMMAND? SH A PRNT:
FREE DISK SPACE:   500 QUANTA

  COMMAND? TIME;
DATE 30/01/79
TIME 22*12*51
  COMMAND? TIME:
DATE 30/01/79
TIME 22*12*56
```

Рис. 4

Протокол измерения времени выполнения команды SHOW ACCOUNT

COMMAND	USER	TERMINAL	ERRCOD	RECORDS	ARRIVALTIME	RESPONSETIME
4	0	[illegible]	0	0	843040	28
13	2	[illegible]	0	0	843135	19
6	4	[illegible]	114	0	843202	24
19	4	[illegible]	0	27	843226	169
7	4	[illegible]	0	25072	843524	5
4	0	[illegible]	0	0	843697	28
13	6	[illegible]	0	3000	843720	18
13	7	[illegible]	0	1000	843727	18
13	6	[illegible]	0	0	843824	3
8	3	[illegible]	113	0	844101	25
13	3	[illegible]	0	0	844247	10
13	3	[illegible]	0	0	844306	1
13	3	[illegible]	0	0	844355	1
13	3	[illegible]	0	0	844405	1
13	3	[illegible]	0	0	844456	0
13	3	[illegible]	0	0	844506	1
13	1	[illegible]	0	0	844521	1
13	7	[illegible]	0	1000	844522	3
13	3	[illegible]	0	0	844557	1
13	1	[illegible]	0	0	844593	3
13	1	[illegible]	0	0	844617	0
13	1	[illegible]	0	0	844843	2
4	0	[illegible]	0	0	844859	15
13	2	[illegible]	0	0	845011	12
13	1	[illegible]	0	0	845026	1
13	1	[illegible]	0	0	845208	2
13	3	[illegible]	0	0	845221	1
6	1	[illegible]	114	0	845318	13
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

21	1	1	0	0	1129230	76
21	1	1	0	0	1129782	66
21	1	1	0	0	1130086	67
21	1	1	0	0	1130760	70
21	1	1	0	0	1130953	69
21	1	1	0	0	1131186	34
21	1	1	0	0	1131490	70
21	1	1	0	0	1132079	72
21	1	1	0	0	1132315	56
21	1	1	0	0	1132546	58
21	1	1	0	0	1133064	56
21	1	1	0	0	1133343	72
21	1	1	0	0	1133686	58
21	1	1	0	0	1134294	35
21	1	1	0	0	1134601	68
21	1	1	0	0	1134791	33
5	1	1	0	25072	1134915	2
5	1	1	0	25072	1134972	1

Рис. 5

Результаты измерения времени выполнения команды SHOW ACCOUNT

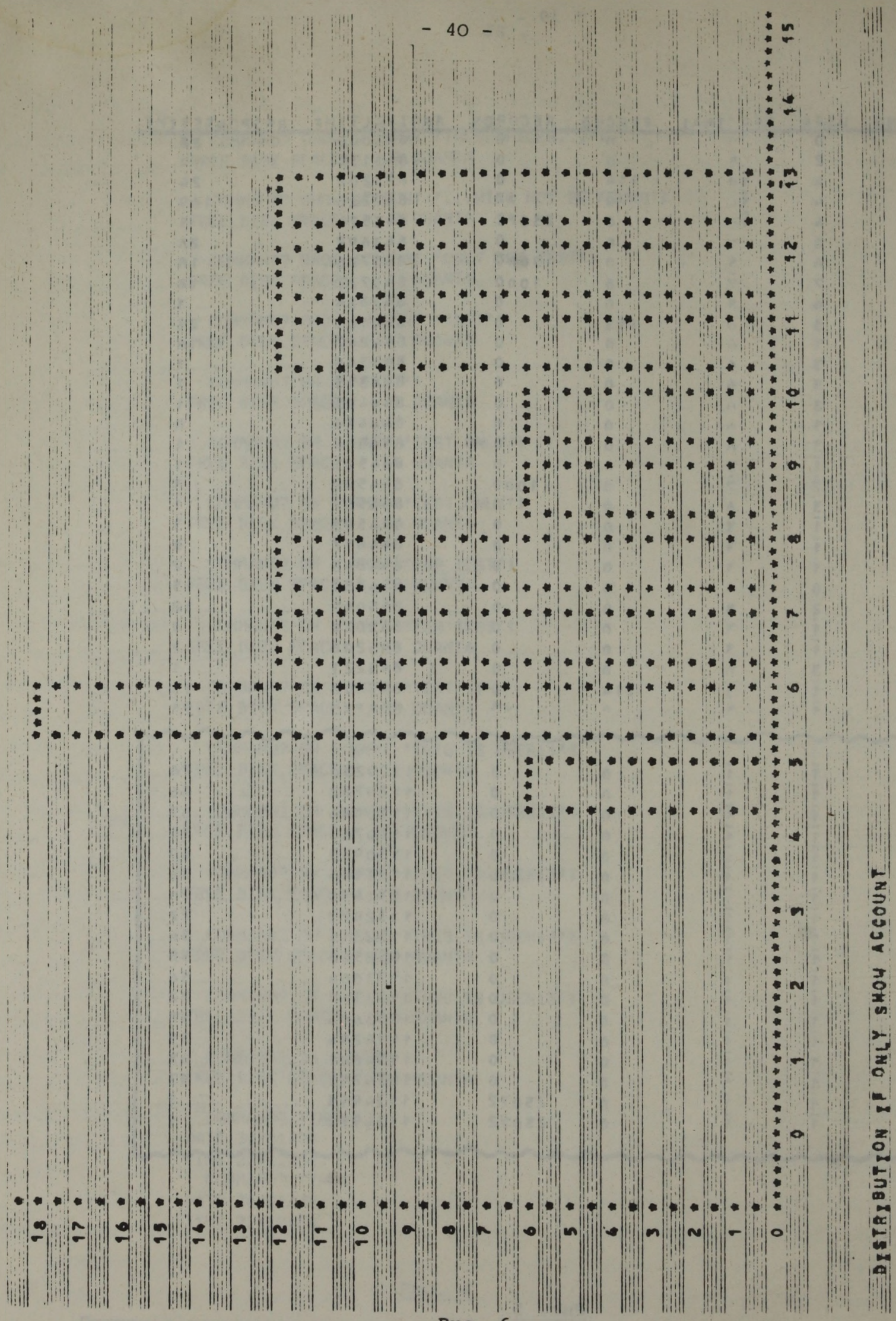


Рис. 6

Результат моделирования времени выполнения команд SHOW ACCOUNT

обусловлены тем, что в модели было предположено больше число файлов на пакетах дисков.

Для отладки модели нужно проделать много таких тестов в краевых условиях. В ходе такого сопоставления можно выявить логические ошибки модели.

После проведения такой отладки составляющих частей модели можно сверять общую статистику измерений с полной работой модели. На рис. 7 приведен результат измерения общих времен ответа для всех команд, а на рис. 8 виден график, соответствующий моделированию. Математическое ожидание общего /для всех команд/ времени ответа, рассчитанное по формуле

$$E(x) = \frac{\sum_{i=1}^{i=N} x_i}{N}$$

в результатах измерении: E(x) = 4,2 секунды
в результатах модлеирования: E (X) =4,5 секунды
Стандартное отклонение, расчитанное по формуле

$$\sigma = \sqrt{\frac{\sum_{i=1}^{i=N} \left[x_i - E(x) \right]^2}{N - 1}}$$

в результатах измерения: σ = 8,1 сек
в результатх моделирования: σ = 5,8 сек
Стандартное отклонение измерении выше, потому что при моделировании не учитывалось время восстановления от ошибок диска T_B /раздел 4./.

8. Использование модели для развития системы

Проверенная модель может быть использована для дальнейшего развития системы. Для расширения машины ЕС-1040 будут покупать накопители на магнитных дисках с большей емкостью и с более быстрым доступом. Стоит вопрос: насколько будут улучшены времена ответа редактора CEDRUS. На рис. 9. приведен результат моделирования работы программы редактора при предположении, что доступ к записи на дисках, два раза меньше. Математическое ожидание общего /для всех команд/ времени ответа E(X) = 0,98 сек. При увеличинеии нагрузки на систему (на 25 %). более частой подаче буферов связи это значение увеличивается лишь незначительно E(X) =1,19 сек. Таким образом можно ответить на вопрос: сколько терминалов стоит еще купить для системы CEDRUS, если она будет эксплуатироваться с более быстрыми дисками.

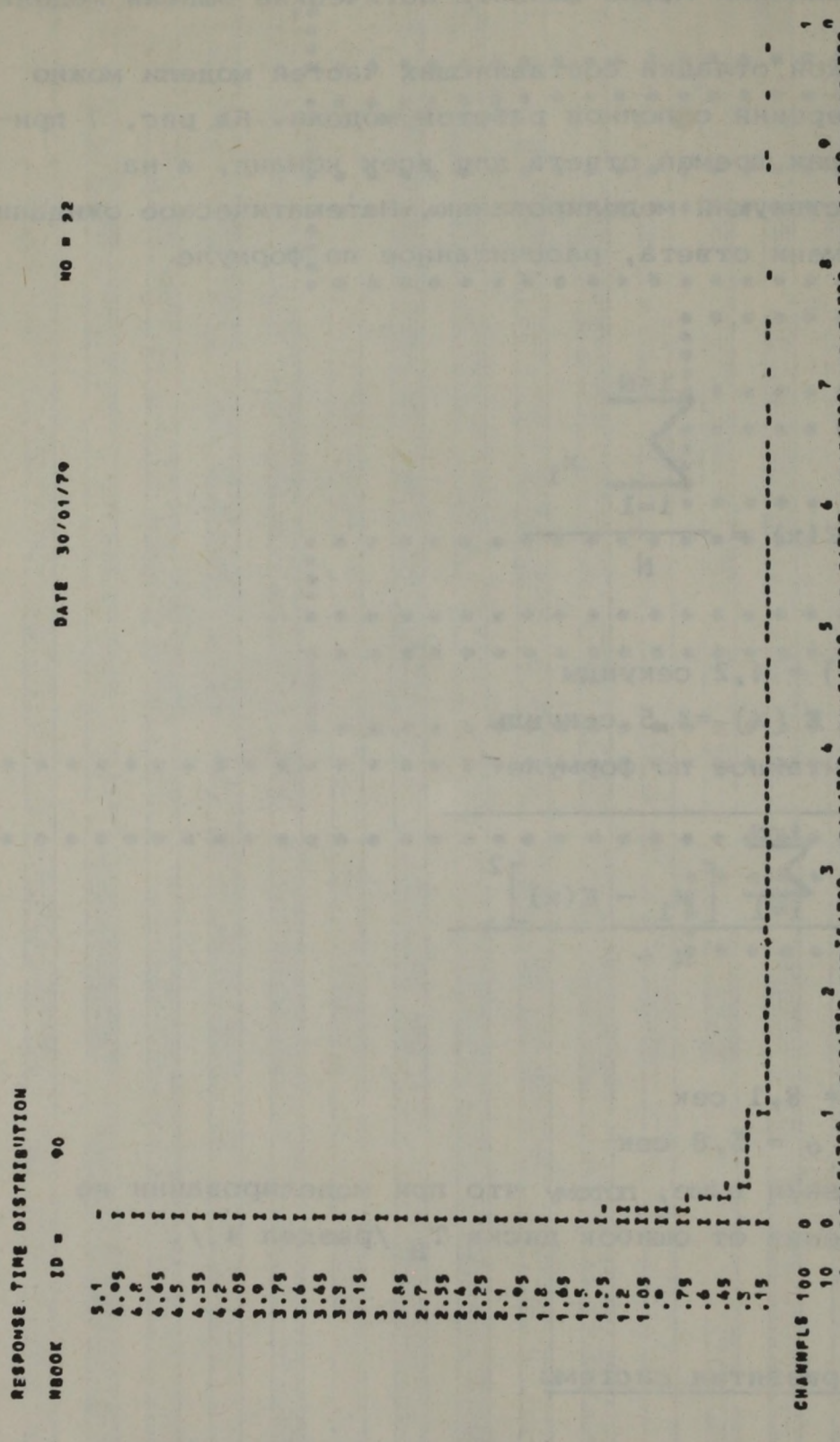


Рис. 7

Распределение времен ответа. Результат измерения

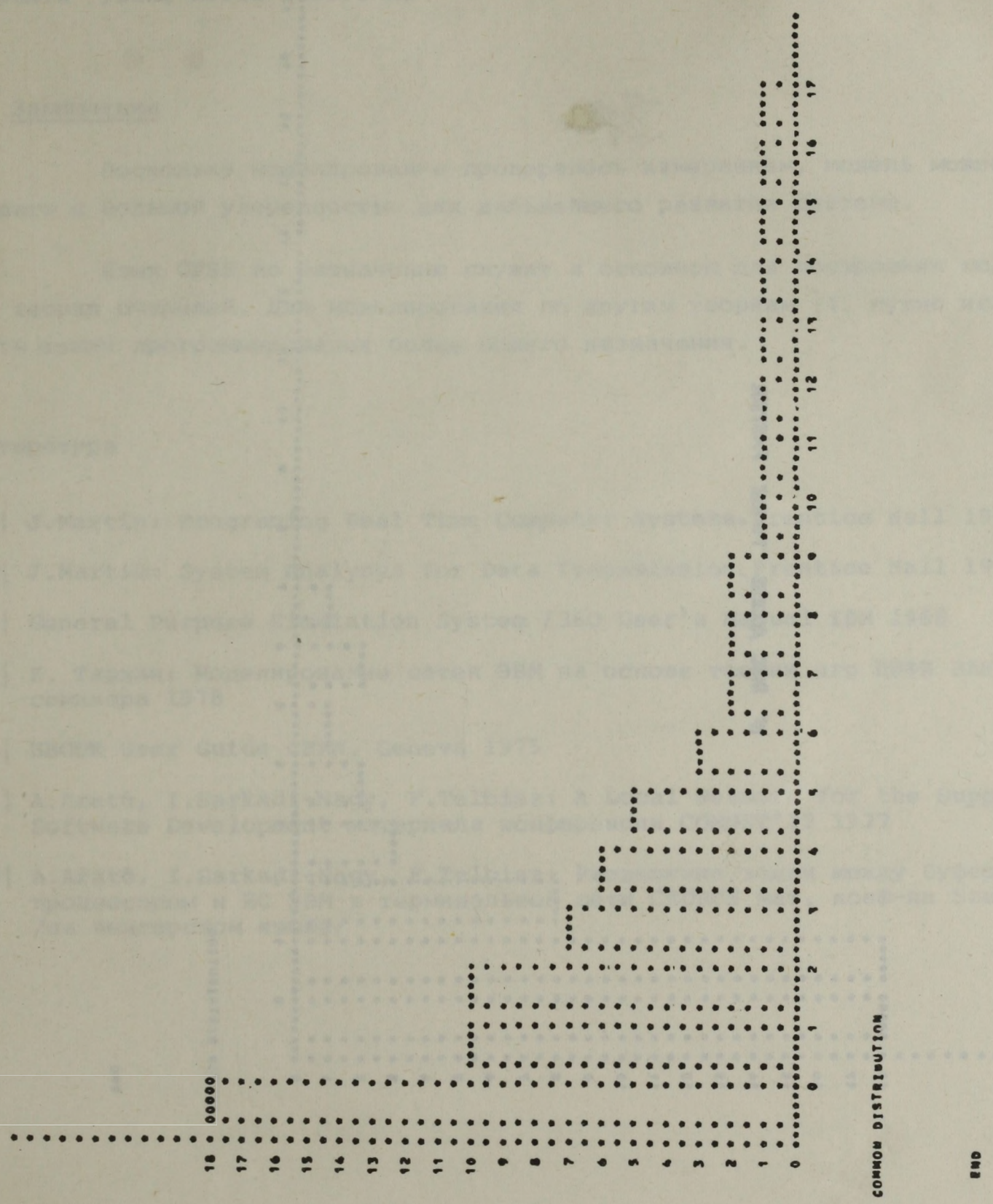


Рис. 8

Распределение времен ответа. Результат моделирования

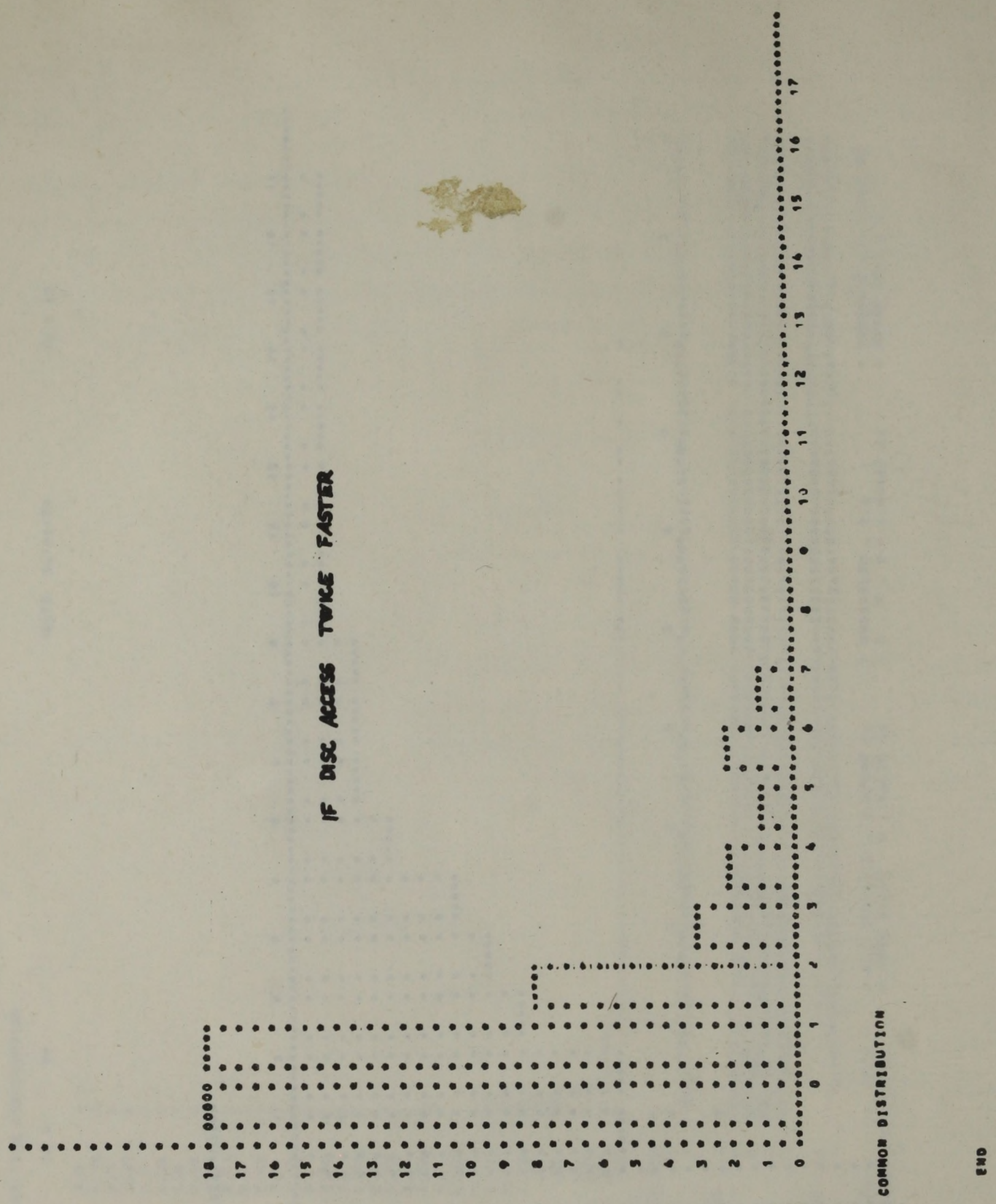


Рис. 9

Распределение времен ответа. Результат моделирования, когда доступ к дискам два раза быстрее

Моедль может быть еще использован для улучшения показателей системы чисто программными средствами. Можно определить те структурные изменения и их параметры, с помощью которых можно изменять времена ответа. Модель помогает выявить "узкие метса" системы.

9. Заключение

Поскольку моделирование проверялось измерением, модель можно использовать с большой уверенностью для дальнейщего развития системы.

Язык GPSS по назначению служит в основном для построения моделей по теории очередей. Для моделирования по другим теориям |4| нужно использовать языки программирования более общего назначения.

Литература

|1| J.Martin: Programing Real Time Computer Systems. Prentice Hall 1971

|2| J.Martim: System Analysys for Data Transmission. Prentice Hall 1972

|3| General Purpose Simulation System /360 User's Manual IBM 1968

|4| К. Тарнаи: Моделирование сетей ЭВМ на основе теории игр ЦИФИ ВАН мат. семинара 1978

|5| HBOOK User Guide CERN, Geneva 1975

|6| A.Arató, I.Sarkadi-Nagy, F.Telbisz: A Local Network for the Support of Software Development материала конференции COMNET'77 1977

|7| A.Arató, I.Sarkadi-Nagy, F.Telbisz: Разделение задач между буферным процессором и ЕС ЭВМ в терминальной сети CEDRUS мат. конф-ии Szeged 1977 /на венгерском языке/

МАТЕМАТИЧЕСКИЕ ВОПРОСЫ ОБРАБОТКИ КАЛИБРОВОЧНЫХ ИЗМЕРЕНИЙ

И.Байла*, Г.А.Ососков**

* Институт измерения и измерительной техники Словацкой Академии Наук, Братислава, ЧССР

**Объединенный институт ядерных исследований, Дубна, СССР

АННОТАЦИЯ

Рассматривается отношение между случайной и детерминистской сторонами задачи калибровки измерительных устройств с прямоугольной системой координат, предназначенных для обработки снимков с трековых камер.

Исследуются математические и методические проблемы поиска калибровочного преобразования, которое определено как наилучшая линейная аппроксимация конечного множества измеренных координат полиномами от двух переменных в L_2 - норме.

Предложен аналитический алгоритм построения аппроксимирующих полиномов, ортогональных на единичном квадрате. Программа ортогонализации, основанная на методе Грамма-Шмидта, написана на языке REDUCE-2, предназначенном для символьных алгебраических вычислений.

Обсуждаются результаты практических расчетов по полученной универсальной калибровочной программе.

ABSTRACT

The relation is considered between random and deterministic features of the calibration problem for measuring devices with rectangular coordinate systems, processing the track chamber slides.

The mathematical and methodological problems of the calibration transform determination are examined. The calibration transformation is defined as the best least square approximation of finite measured coordinate set by bivariate polynomials.

The analytical algorithm for constructing approximating polynomials orthogonal on the unit interval is given. The orthogonalization program, based on Gramm-Schmidt method, is written in REDUCE -2 language designed for algebraic computations.

Applied calculation results of the universal calibration program are discussed.

I. Введение

Одной из важнейших задач физики высоких энергий является проблема массовой обработки потоков экспериментальной информации, получаемых путем автоматического измерения стереофотографий, на которых сняты события, происходящие в трековых камерах. Измерение снимков производится с помощью быстрого просмотра (сканирования) изображения световым пятном с последующей оцифровкой получаемого сигнала.

Таким образом основными данными, выдаваемыми сканирующими автоматами и просмотрово-измерительными устройствами, обрабатывающими фильмовую информацию с трековых камер, являются упорядоченные пары чисел. Эти пары должны представлять точки измеряемого снимка в некоторой унифицированной системе декартовых координат. Невозможность получить такое предоставление непосредственным путем является общей чертой упомянутых измерительных устройств и приводит к задаче их калибровки, решение которой позволяет правильно интерпретировать данные измерений, а также определить точностные характеристики измерительных приборов.

Актуальность вопросов калибровки и разработки соответствующего программного обеспечения породила немало работ (см., например, /I-3/), однако в них мало внимания уделялось математической стороне дела и вычислительным особенностям алгоритмов и программ калибровки.

В настоящей работе, являющейся обобщением работ /4,5/, анализируется связь случайной и детерминистской стороны задачи калибровки и рассматриваются вычислительные проблемы поиска наилучшей аппроксимации конечного множества измеренных координат полиномом двух переменных в L_2 - норме. Удобным математическим аппаратом для этой цели являются системы ортогональных двумерных полиномов. Описанный ниже метод построения таких систем основан на использовании алгоритмического языка программирования REDUCE-2, предназначенного для символьных алгебраических вычислений.

2. Постановка задачи

Будем представлять себе снимок, как двумерный набор точек, подлежащих измерению. Тогда с каждой точкой A на плоскости снимка можно связать две пары чисел:

- её декартовы координаты в этой плоскости;
- пару целых чисел x_m , y_m , получаемых с отчетных устройств нашего прибора в результате акта измерения точки A .

Вследствие случайных и систематических дисторсий прибора и ошибок измерения снимка может оказаться, что не существует такого линейного преобразования L , чтобы для всех m было $L(x_m, y_m) = (x, y)$.

Возникает задача: разработать для конкретного измерительного прибора методику обработки данных измерения, позволяющую верное толкование полученных результатов. Основой такой методики является калибровочное измерение, при котором в приборе измеряется специальная эталонная (калибровочная) решетка, состоящая из конечного набора крестов. Так как декартовые координаты этих крестов заранее определены с высокой точностью ($\sim 1\mu$), соответствующую систему декартовых координат отождествляем с идеальной.

Обозначим через x_i , y_i идеальные координаты i -го ($i = \overline{1,n}$) креста, а через x_{mij} , y_{mij} - измеренные значения координат ($j = \overline{1,l}$ - индекс текущего измерения).

Далее, пусть через Δx_{ij} , Δy_{ij} обозначено случайное отклонение, а через s_i^x , s_i^y - систематическое отклонение измеренных от идеальных значений. Тогда для описания соотношения идеальных и измеренных значений можно принять следующую модель:

$$x_{mij} = x_i + \Delta x_{ij} + s_i^x \quad , \qquad (1)$$

$$y_{mij} = y_i + \Delta y_{ij} + s_i^y \quad , \qquad (2)$$

где относительно случайных и систематических отклонений делаются следующие исходные допущения:

1) величины s_i^x , s_i^y ($i = \overline{1,n}$) являются значениями некоторых функций $f_x(x, y; t)$, $f_y(x, y; t)$ от двух переменных в точках плоскости (соответствующих центрам крестов), определенные и непрерывные на некотором двумерном интервале и, помимо того, зависящие от параметра - времени t ;

2) случайные отклонения (погрешности) Δx_{ij} , Δy_{ij} независимо от метода определения центров крестов являются независимы-

ми нормально распределенными случайными величинами с нулевыми средними и среднеквадратичными значениями σ_{x_i} и σ_{y_i} соответственно, т.е. распределенными по законам $N(0, \sigma_{x_i})$, $N(0, \sigma_{y_i})$.

Необходимой предпосылкой применимости метода калибровочных измерений является установление периода стабильности калибруемого прибора относительно систематических отклонений. Для этой цели предполагается:

- установить оценку периода стабильности на основе нескольких последовательных циклов измерений, проведенных на рассматриваемой измерительной установке;

- проверить полученную оценку при помощи критерия для проверки статистической гипотезы об отсутствии систематического сдвига в наблюдениях.

Излагая этот критерий, будем рассматривать только случай наблюдений x_{mij}, опуская для краткости индекс m.

Пусть калибровочная решетка измерена последовательно l раз. Каждое наблюдение $\tilde{x}_{ij}$ можно считать результатом измерения неизвестного значения величины $\tilde{x}_i = x_i + s_i^x$. Тогда по предположению 2) принятой модели для каждого измеряемого креста имеем $E(\tilde{x}_{ij}) = \tilde{x}_i = x_i + s_i^x$

Для проверки гипотезы о стабильности $\tilde{x}_i$ вводятся статистики /6/:

$$q_i^2 = \frac{1}{2(l-1)} \sum_{j=1}^{l-1} (\tilde{x}_{i,j+1} - \tilde{x}_{ij})^2 \quad , \quad (3)$$

$$s_i^2 = \frac{1}{l-1} \sum_{j=1}^{l} (\tilde{x}_{ij} - \bar{x}_i)^2 \quad , \quad (4)$$

где $\bar{x}_i = \frac{1}{l} \sum_{j=1}^{l} \tilde{x}_{ij}$ и определяется частное $r_i = q_i^2 / s_i^2$.

Если в течение измерений присутствует систематический сдвиг $E(\tilde{x}_{ij})$, то следует ожидать, что s_i^2 будет много больше, чем q_i^2. Пользуясь таблицами квантилей r_{ip} порядка p распределения случайной величины r_i /6/ можно при заданном порядке p (обычно $p = 0{,}05$) сравнить значения r_{ip} и r_i. Если окажется, что $r_i < r_{ip}$, то считаем, что в наблюдениях $\tilde{x}_{ij}$ в i-м кресте существует сдвиг, объясняемый нарушением гипотезы стабильности $E(\tilde{x}_{ij})$ по времени.

В случае положительных результатов проверки можно считать,

что систематические отклонения измерений f_x, f_y не зависят от времени t .

Самое простое предположение относительно аналитического выражения этих функций - их линейность, т.е. наличие таких чисел a_{11}, a_{12}, a_{21}, a_{22}, x_o, y_o, для которых имеют место равенства

$$f_x(x_i, y_i) = s_i^x = x_i - (a_{11}x_i + a_{12}y_i + x_o),$$

$$f_y(x_i, y_i) = s_i^y = y_i - (a_{21}x_i + a_{22}y_i + y_o) \qquad \forall i = \overline{1,n},$$

где по-прежнему x_i, y_i - значения идеальных координат эталонных крестов.

Однако в случае сканирующих автоматов, особенно с электронно-лучевым сканированием, как показывает первый же взгляд на любое изображение результатов измерений, они уже не допускают возможности такой простой интерпретации из-за появления нелинейных, обычно бочкообразных или подушкообразных искажений. Возникает необходимость найти такое преобразование $\mathcal{A} : E_2 \longrightarrow E_2$, которое на основе обработки данных измерения калибровочной решетки каждой точке $X_m = (x_m, y_m)$ сопоставит точку $X^* = \mathcal{A}(X_m)$, причем расстояние $\rho(X^*, X)$ должно быть минимальным в смысле некоторого выбранного критерия (здесь $X = (x,y)$; x,y - идеальные координаты).

Обозначим конечные множества идеальных координат x_i и y_i через $\mathcal{M}_x$ и $\mathcal{M}_y$, а через $P(x, y; A_x)$ - линейную аппроксимирующую функцию вида

$$P(x,y;A_x) = \sum_{j=1}^{m} a_{xj}\varphi_j(x,y) \qquad , \qquad (5)$$

где a_{xj} - элементы неизвестного вещественного вектора A_x, φ_j - некоторые линейно независимые непрерывные функции от двух переменных.

Пусть w_{xi} - система весовых значений по координате x, причем $w_{xi} > 0$, $i = \overline{1,n}$.

Выражением

$$L_p(g) = \left[\sum_{i=1}^{n} |g(x_i, y_i)|^p\right]^{1/p}, \quad p \geq 1,$$

определяется дискретная L_p - норма функции $g(x, y)$,

заданной в конечном числе точек (x_i , y_i). Точно также вводятся обозначения в случае функции $Q(x, y ; A_y)$ и весов w_{yi} .

Тогда задачу калибровки можно свести к двум отдельным проблемам теории линейной аппроксимации функций, относящимся к двум множествам координат $\mathcal{M}_x$, $\mathcal{M}_y$:

- требуется найти наилучшую линейную аппроксимирующую функцию $P(x, y ; A_x)$ (или $Q(x, y ; A_y)$) конечного множества $\mathcal{M}_x$ (или $\mathcal{M}_y$) координат x_i (или y_i), которые считаем значениями какой-нибудь функции h_x (или h_y) от двух переменных, заданной в конечном числе точек $Z_i = (x_{mi}, y_{mi})$, в дискретной L_2 - норме с весами w_{xi} (или w_{yi}).

Следовательно, надо искать минимум функции расстояния:

$$L_2(w_x, P-h_x)^2 = \sum_{i=1}^{n} | P(x_{mi}, y_{mi}; A_x) - x_i |^2 \cdot w_{xi} \qquad (6)$$

(или $L_2(w_y, Q-h_y)^2 = \sum_{i=1}^{n} | Q(x_{mi}, y_{mi}; A_y) - y_i |^2 \cdot w_{yi}$).

Решая эти две проблемы, т.е. находя векторы параметров A_x^*, A_y^* , можем каждой измеренной точке снимка $X_m = (x_m , y_m)$ сопоставить точку

$$X_m^* = (x_m^*, y_m^*) = (P(x_m, y_m; A_x^*), Q(x_m, y_m; A_y^*)),$$

и тем самым определить калибровочное преобразование

$$\mathcal{H} : (x_m, y_m) \longrightarrow (x_m^*, y_m^*).$$

Требование относительной простоты калибровочного преобразования и его представления единственной формулой для всех измеренных данных приводит к идее выбора аппроксимирующей функции в виде полинома от двух переменных:

$$P_m(x, y; A_x) = \sum_{j=0}^{m} \sum_{i=0}^{j} a_{xq} x^{j-i} y^i \qquad , \quad (7)$$

где $q = \left[\frac{j(j+1)}{2} + i\right]$, m - максимальная степень полиномов.

Программы обработки снимков с трековых камер требуют кроме прямого также и обратное преобразование координат, определяющее переход от идеальных к измеренным координатам. Поскольку, нелинейность прямого преобразования не позволяет получить коэффициенты обратного преобразования с помощью обращения его матрицы, об-

ратное преобразование определяется опять-таки при помощи полиномов двух переменных, т.е. функциями $P^{-1}(x, y ; A'_x)$, $Q^{-1}(x, y ; A'_y)$, заданным теперь в точках (x_i, y_i) идеальной решетки и линейно-аппроксимирующих:

- в случае P^{-1}, конечное множество $\mathcal{M}'_x$ измеренных координат x_{mi} ;
- в случае Q^{-1}, конечное множество $\mathcal{M}'_y$ измеренных координат y_{mi} .

3. Вычислительные аспекты задачи калибровки

В задачах аппроксимации непрерывной функции линейно аппроксимирующей функцией в дискретной L_2 - норме возникают вычислительные проблемы, связанные с плохой обусловленностью матрицы системы нормальных уравнений. Один из возможных способов решения этой проблемы, широко распространенный на практике, заключается в применении для решения системы линейных уравнений программ с двойной точностью [I,3].

Однако, можно выбрать другой, более эффективный подход, состоящий в преобразовании матрицы системы к диагональному виду с использованием для аппроксимации линейной комбинации функций, ортогональных на конечном множестве данных точек. Таким образом, в нашем случае возникает задача ортогонализации произвольной системы $\{\varphi_n\}$ линейно независимых полиномов.

Любой полином ортогональной системы $\{\psi_n\}$ можно искать в виде линейной комбинации исходных полиномов φ_n :

$$\psi_n = \sum_{i=1}^{n} c_{in} \varphi_i \qquad (c_{nn} \neq 0)$$

с неизвестными коэффициентами c_{in} , ($i = 1,2,\ldots,n$).

Если множество точек, на котором должны быть ортогональны вычисляемые полиномы, фиксировано, то вычисления соответствующих коэффициентов необходимо провести только один раз. Из этого следует, что с вычислительной точки зрения основной выигрыш использования таких ортогональных полиномов получается тогда, когда аппроксимирующие функции определены при фиксированных значениях своих аргументов.

В случае определения полиномов от двух переменных на множестве точек (x_i, y_i), которые не фиксированы, но размещены почти симметрично относительно центра системы декартовых координат, естественно возникает мысль использовать полиномы от двух пере-

менных, ортогональные в непрерывном смысле, так как для устранения вычислительных проблем, возникающих из-за ошибок округления при обращении плохо обусловленных матриц, достаточно /7/ вместо диагональной матрицы использовать приблизительно диагональную.

Следует заметить важный с вычислительной точки зрения факт различия множеств точек Z_i, в которых заданы полиномы P и Q в каждом отдельном калибровочном измерении при прямом преобразовании. Для обратного преобразования это множество точек фиксировано, поскольку состоит из точек идеальной решетки. Тогда оказывается более выгодным определить калибровочные преобразования при помощи двух разных систем полиномов:

I) системы полиномов, ортогональных на единичном квадрате $Q_2 = \langle -1,1 \rangle \times \langle -1,1 \rangle$ для прямого преобразования;

2) системы полиномов, ортогональных на множестве узлов идеальной решетки для обратного преобразования.

4. Полиномы от двух переменных, ортогональные на единичном квадрате $Q_2 = \langle -1,1 \rangle \times \langle -1,1 \rangle$

Пусть $\{\varphi_n\}$ – система полиномов от двух переменных x, y вида $x^\alpha y^\beta$, записанных в лексикографическом порядке:

$$\{\varphi_n\} = \{1, x, y, x^2, xy, y^2, \dots\} \quad , \qquad (9)$$

которые определены на квадрате Q_2.

Скалярное произведение двух полиномов будем задавать как

$$(\varphi_j, \varphi_k) = \iint_{-1-1}^{11} \varphi_j(x,y)\,\varphi_k(x,y)\,dx\,dy \, . \qquad (10)$$

Поскольку система полиномов (9) линейно независима, можно на её основе построить систему $\{\psi_n\}$ полиномов, ортогональных на квадрате Q_2 относительно скалярного произведения (10), используя метод ортогонализации Грамма-Шмидта /7/:

$$\psi_n = \varphi_n - \sum_{j=1}^{n-1} \frac{(\psi_j, \varphi_n)}{(\psi_j, \psi_j)} \cdot \psi_j \, . \qquad (11)$$

Следует отметить, что

I) эта рекурентная формула решает задачу ортогонализации лишь теоретически, так как её применение для практического вычисления ортогональных полиномов высших степеней оказывается слишком громоздким;

2) рекурентный характер формулы (II) в связи с требованием

вычислять определенные интегралы (ψ_j , φ_n) и (ψ_j , ψ_j) не дает возможности реализовать вычислительный алгоритм при помощи обычных систем программирования.

Однако, именно для решения такой проблемы последовательного вычисления аналитических формул и их использования в одной и той же программе разработаны языки и системы символьного программирования алгебраических вычислений, такие, как REDUCE - 2 [8].

Мы не будем подробно описывать алгоритм построения аналитических формул для вычисления полиномов от двух переменных, ортогональных на квадрате Q_2 , так как он достаточно ясен из прилагаемой программы, написанной на языке REDUCE - 2.

INPUT

```
N := 50
   %NUMBER OF ORTHOGONAL POLYNOMIALS REQUIRED;

ON DIV;

ARRAY PHI(N),PSI(N);
   %ARRAYS FOR STORING PHI'S AND PSI'S;

%SET UP INITIAL PHI ARRAY;

INTEGER PROCEDURE FLOOR K;
   %FINDS LARGEST INTEGER M SUCH THAT M**2+M<2*K;
   BEGIN INTEGER M;
      K := 2*K;
      M := 1;
      WHILE M**2+M<K DO M := M+1;
      RETURN M
   END;


FOR I := 1:N DO
   <<N1 := FLOOR I; N2 := I-N1*(N1-1)/2;
   PHI(I) := X**(N1-N2)*Y**(N2-1)>>;

COMMENT NOW SET UP INTEGRATION METHOD;
COMMENT THIS CAN EITHER BE BY PATTERNS, AS BELOW,
OR A MORE GENERAL METHOD;

OPERATOR INT;

LINEAR INT;

FOR ALL I,X LET INT(1,X)=X,
                INT(X,X)=X**2/2,
                INT(X**I,X)=X**(I+1)/(I+1);


COMMENT NOW DEFINE DEFINITE INTEGRATION REGION;

ALGEBRAIC PROCEDURE DINT(U);
   BEGIN SCALAR Z;
      Z := INT(U,X);
      Z := SUB(X=1,Z)-SUB(X=-1,Z);
      Z := INT(Z,Y);
      RETURN SUB(Y=1,Z)-SUB(Y=-1,Z)
   END;


COMMENT NOW COMPUTE PSI'S;

FOR I:= 1:N DO
   WRITE PSI(I) := PHI(I)-FOR J :=1:I-1 SUM DINT(PSI(J)
                   *PHI(I))*PSI(J)/DINT(PSI(J)**2);
```

OUTPUT

```
PSI(1) := 1
PSI(2) := X
PSI(3) := Y
PSI(4) := X**2 - 1/3
PSI(5) := X*Y
PSI(6) := Y**2 - 1/3
PSI(7) := X**3 - 3/5*X
PSI(8) := X**2*Y - 1/3*Y
PSI(9) := X*Y**2 - 1/3*X
PSI(10) := Y**3 - 3/5*Y
PSI(11) := X**4 - 6/7*X**2 + 3/35
PSI(12) := X**3*Y - 3/5*X*Y
PSI(13) := X**2*Y**2 - 1/3*X**2 - 1/3*Y**2 + 1/9
PSI(14) := X*Y**3 - 3/5*X*Y
PSI(15) := Y**4 - 6/7*Y**2 + 3/35
PSI(16) := X**5 - 10/9*X**3 + 5/21*X
PSI(17) := X**4*Y - 6/7*X**2*Y + 3/35*Y
PSI(18) := X**3*Y**2 - 1/3*X**3 - 3/5*X*Y**2 + 1/5*X
PSI(19) := X**2*Y**3 - 3/5*X**2*Y - 1/3*Y**3 + 1/5*Y
PSI(20) := X*Y**4 - 6/7*X*Y**2 + 3/35*X
PSI(21) := Y**5 - 10/9*Y**3 + 5/21*Y
```

Правильность полученных формул можно проверить перемножая P_n-одномерные полиномы Лежандра соответствующих степеней /9/, т.к. для одномерных полиномов $\Psi_n^1 = [2^n(n!)^2/(2n)!]P_n$.

5. Программная реализация

На основе проведенных исследований были разработаны универсальные программы GENCAL и ORCAL. Эти программы предназначены для вычисления коэффициентов калибровочных преобразований для измерительной аппаратуры, работающей в декартовой системе координат. В качестве критериев эффективности найденных преобразований взята максимальная остаточная ошибка R_{max} (R'_{max}), среднее значение остаточных ошибок $\bar{R}$ ($\bar{R}'$) по полю и их среднеквадратическое значение $\sigma_{\bar{R}}$ ($\sigma_{\bar{R}'}$). Программы управляются с помощью параметров, определяющих выбор числа и расположения крестов в калибровочной решетке, точность, с которой задается центр каждого из крестов, вид и степень аппроксимирующих полиномов.

Программы реализованы на языке ФОРТРАН для ЭВМ СДС-6500 ОИЯИ. Результатом работы этих программ кроме массивов коэффициентов для прямого и обратного преобразования являются таблицы и гистограммы остатков, значения точностных параметров R_{max}, $\bar{R}$, $\sigma_{\bar{R}}$.

Печать, которой оснащены эти программы, увеличивает возможности использования полученных количественных результатов в качестве широкого проверочного средства для контроля стабильности калибруемого измерительного устройства.

На обширном практическом материале, относящимся к приборам различного типа (измерительные столы "CAMET", автоматы на электроннолучевых трубках " ERASME ", "АЭЛТ 2/I60", измерительная видиконная система), было проведено большое количество калибровочных расчетов. Целью их явилась необходимость проверки правильности работы программ путем сравнения с уже имевшимися расчетами, оптимизация степени полиномов, исследование качества аппроксимации между узлами решетки, а также расширение возможностей программ, т.е. универсализация их для применения к широкому кругу различных приборов.

Результаты расчетов (бóльшая их часть приведена в /4/) показывают, что

- для измерительных столов типа CAMET достаточно взять в качестве калибровочного преобразования полином первой степени, так как увеличение точности с ростом степени аппроксимирующего полинома невелико, в то время как затраты времени для счета растут значительно;

- для достаточно точной корректировки систематических отклонений в сканирующих автоматах типа ERASME и АЭЛТ-2/I60 необходимо использовать полиномы 5-ой степени. Повышение степени полиномов приводит к медленному улучшению аппроксимации в узлах эталонной решетки, в то время как между узлами качество аппроксимации ухудшается. Для степеней выше 5 значения точностных параметров становятся для задачи калибровки недопустимо большими.

Таким образом выбор конкретной степени I÷5 зависит от вида калибруемого прибора (искажений) и не зависит от хода калибровочных измерений, т.е. степень полиномов достаточно установить для данного прибора один раз при первоначальной калибровке.

Заключение

I. Сравнение результатов расчетов калибровочных параметров церновской установки ERASME , проведенных по одним и тем же данным по церновской калибровочной программе и по GENCAL и ORCAL , показало совпадение результатов с точностью, обеспе-

чиваемой отсчетной системой прибора.

2. Полученными в § 4,5 формулами ортогональных полиномов можно пользоваться не только в задачах калибровки, но также в любых задачах аппроксимации непрерывной функции f от двух переменных на единичном квадрате в L_2 - норме.

Аналитическая программа ортогонализации наряду с получением формул, определяющих полиномы, ортогональные на единичном квадрате Q_2 в явном виде, дает идею аналитического вычисления таких полиномов также и в случаях других исходных систем функций или другого, чем в (II), скалярного произведения.

3. Эксплуатация программ GENCAL и ORCAL показала их универсальность, обеспечиваемую возможностью гибкой перестройки программ и достаточно высокое быстродействие (4 сек на обработку одного сеанса калибровки, по сравнению, например, с I5 сек по программе /3/).

4. Программы GENCAL и ORCAL включены в состав математического обеспечения автомата "АЭЛТ-2/I60 и находятся в постоянной эксплуатации с июня I978 года. Это матобеспечение дополнено программой TEST для проверки стабильности автомата.

Литература

I. Аникеев В.Б. и др. ИФВЭ, ОМВТ 75-9I, Серпухов, I975.
2. Klein F., Ströbele H. A Fortran Program for PEPR Calibration using Chebyshew Polynomials. Proceedings of International Conference on Data Handling Systems in High-energy Physics. Cambridge 1970. CERN 70-21.
3. Карлов А.А., Сенченко В.А. ОИЯИ, I0-III55, Дубна, I977.
4. Байла И., Ососков Г.А. ОИЯИ, PI0-II834, Дубна, I978.
5. Байла И., Ососков Г.А., Хэрн А.К. ОИЯИ, PI0-II944, Дубна, I978.
6. Линник Ю.В. Метод наименьших квадратов и основы теории обработки наблюдений. ГИФМЛ, Москва, I962.
7. Rice J.R. The Approximation of Functions. Vol.1.Linear Theory. Addison-Wesley publishing Company, INC. London 1964.
8. Hearn A.C. REDUCE-2 user's manual. University of Utah, Salt Lake City,1973.
9. Градштейн И.С., Рыжик И.М. Таблицы интегралов,сумм,рядов и произведений. ГИФМЛ.Москва,I962.

НЕКОТОРЫЕ МЕТОДЫ ПРИБЛИЖЕННОГО РЕШЕНИЯ УРАВНЕНИЙ ТИПА ЛОУ

Е.П.Жидков, М.Нгуен, Б.Н.Хоромский
Объединенный институт ядерных исследований, Дубна

АННОТАЦИЯ

Рассмотрены некоторые математические вопросы, связанные с приближенным решением системы нелинейных сингулярных уравнений типа Лоу. Установлены условия существования и единственности решений этих уравнений.

ABSTRACT

We consider some mathematical questions concerning approximate solutions of nonlinear singular integral Low equations systems.

Conditions of existence and uniqueness of solutions are obtained.

§ I. Введение

В настоящей работе рассматривается нелинейное уравнение Лоу, которое начиная с работы /I/ является объектом всестороннего изучения /2,3,4/. Достаточно подробный обзор полученных в этом направлении результатов представлен, например, в /4/.

Здесь рассмотрим некоторые математические вопросы, связанные с приближенным решением этого уравнения.

Возможны три основные формулировки задачи, связанной с уравнением Лоу: сингулярное интегральное уравнение, краевая задача для аналитической вектор-функции, система разностных уравнений. В первой формулировке искомая функция $h_\alpha(w)$, $\alpha = 1, \dots, N$ комплексной переменной w удовлетворяет системе нелинейных сингулярных интегральных уравнений /I/

$$h_\alpha(w) = \frac{\lambda_\alpha}{w} + \frac{1}{\pi}\int_1^\infty dw' f(w') \left(\frac{|h_\alpha(w')|^2}{w'-w} + \sum_{\beta=1}^{N} \frac{A_{\alpha\beta}|h_\beta(w')|^2}{w'+w} \right), \quad \alpha = 1, \dots, N \qquad (I.1)$$

где числа λ_α и матрица $A = \{A_{\alpha\beta}\}$ - заданы. Эта система может быть сведена при помощи формул Сохоцкого-Племеля к уравнению для граничных значений $h_\alpha(t)$, $t = 1/w$ на верхнем берегу разреза $w \in [1, \infty)$ /5,6/

$$h_\alpha(t) = \lambda_\alpha t + i\rho(t)|h_\alpha(t)|^2 -$$

$$- \frac{t}{\pi}\int_0^1 \frac{\rho(\tau)}{\tau} \frac{|h_\alpha(\tau)|^2}{\tau - t} d\tau + \sum_{\beta=1}^{N} A_{\alpha\beta} \frac{t}{\pi}\int_0^1 \frac{\rho(\tau)}{\tau} \frac{|h_\beta(\tau)|^2}{\tau + t} d\tau . \qquad (I.2)$$

Системе (I.1) соответствует следующая краевая задача /4,7/: найти аналитическую внутри единичной окружности C_0 вектор-функцию $h_\alpha(z)$, $\alpha=1,\dots,N$; $w = 2z/(1+z^2)$, удовлетворяющую условиям

I. $h_\alpha(z)$ имеют непрерывные на C_0 граничные значения.

2. $\overline{h_\alpha(z)} = h_\alpha(\bar z)$ - условие действительности (черта означает операцию сопряжения).

3. $\operatorname{Im} h_\alpha(z) = F(\varphi)\,|h_\alpha(z)|^2$; $z = e^{i\varphi}$, $0 \le \varphi \le \pi/2$ - условие унитарности.

Заданная функция $F(\varphi)$ удовлетворяет условию Гельдера при $\varphi \in [0, \frac{\pi}{2}]$ и $F(0) = F(\pi/2) = 0$.

4. $h_\alpha(-z) = \sum_{\beta=1}^{N} A_{\alpha\beta}\, h_\beta(z)$ есть условие перекрестной симметрии. Заданная вещественная матрица $A = \{A_{\alpha\beta}\}$ удовлетворяет равенству $A^2 = E$, где E - единичная матрица.

5. $h_\alpha(z)$ имеет при $z = 0$ полюс λ_α / z с заданным вычетом λ_α таким, что $\lambda_\alpha = -\sum_{\beta=1}^{N} A_{\alpha\beta}\,\lambda_\beta$.

Условия I-5 представляют нелинейную краевую задачу типа Римана-Гильберта (линейная задача Римана-Гильберта рассматривается, например, в /8,9/) для мероморфного вектора $H(z) = (h_1(z), \dots, h_N(z)) = U + iV$:

$$V(\sigma) - \bar F(\varphi)\, G(U(\sigma), V(\sigma)) = 0;\quad \sigma = e^{i\varphi},\ \varphi \in [-\pi, \pi], \tag{I.3}$$

где

$$\bar F(\varphi) = \begin{cases} F(\varphi), & \varphi \in [0, \pi/2] \\ -F(\varphi), & \varphi \in [-\pi/2, 0] \\ F(\varphi - \pi), & \varphi \in [\pi/2, 3\pi/2], \end{cases}$$

а вектор $G(U, V)$ имеет вид

$$G(U, V) = \begin{cases} U^2 + V^2, & \varphi \in [-\frac{\pi}{2}, \frac{\pi}{2}] \\ A([AU]^2 + [AV]^2), & \varphi \in [\pi/2, 3\pi/2]. \end{cases} \tag{I.4}$$

Здесь и далее символ $X^2=(x_1^2,\dots,x_N^2)$ означает вектор, составленный из квадратов координат вектора $X=(x_1,\dots,x_N)$.

Для матриц кроссинг-симметрии A специального вида

$$\sum_{j=1}^{N} A_{ij} = 1 \quad , \quad i=1,\dots,N \tag{I.5}$$

и в случае, когда $\overline{F(\varphi)}$ есть краевое значение некоторой мероморфной функции $F(z)$, оказалось удобным перейти к изучению элементов S -матрицы /4,I0/ $S(z)=\{S_j(z)\}$, $j=1,\dots,N$

$$S_j(z) = 1 + 2i\,F(z)\,h_j(z), \quad j=1,\dots,N$$

в плоскости униформизующей переменной

$$w = \frac{1}{\pi}\arcsin z \ .$$

При этом условия 2-4 превращаются в следующие

$$\overline{S}(w)=S(\bar{w}) \ ; \quad S(w)\,S(1-w)=1 \ ;$$

$$S(1+w)=[A\,S(w)]^{-1}. \tag{I.6}$$

Для этих нелинейных разностных уравнений исследован локальный вид решений в окрестности неподвижных точек. Для некоторых трехрядных матриц A ($N=3$) найдены решения с конечным числом полюсов на w -плоскости. Подробные исследования уравнений в форме (I.6) можно найти в /3,I0,II,I2/. Отметим, что в /I2/ громоздкие аналитические выкладки, необходимые для построения решений в окрестности неподвижных точек, производятся на ЭВМ.

Для изучения вопросов существования и единственности решений уравнения Лоу без указанных дополнительных предположений плодотворными оказываются методы функционального анализа, позволяющие изучить это уравнение в наиболее общей формулировке. Изучение дифференциальных свойств нелинейного оператора Лоу проводится в п.2, где, в частности, приводится формула для суммарного индекса производной Фреше . Наиболее подробно изучены малые по модулю решения. Некоторые возможности численной

реализации итерационных методов рассмотрены в п.4, где также исследована точность разностных схем и условия сходимости приближенных решений. В табл. I-3 приведены результаты расчетов ряда модельных задач, а также результаты численных экспериментов по определению верхней границы константы связи λ_α, обеспечивающей существование решения.

Основные утверждения мы приводим в строгой формулировке, однако, их доказательства будут иметь лишь конспективный характер. Более подробные детали выкладок можно найти в /5-7,I3-I6/.

Авторы считают своим долгом выразить благодарность проф. И.П.Недялкову за многочисленные и полезные обсуждения физической стороны проблемы.

§ 2.

Перейдем к рассмотрению краевой задачи (I.3). Определим нормированное пространство L_α непрерывных на C_0 вещественных функций, удовлетворяющих условию Гельдера с показателем $0<\alpha<1$ и нормой

$$\|f\|_{L_\alpha} = \sup_{\sigma\in C_0}|f(\sigma)| + \sup \frac{|f(\sigma_1)-f(\sigma_2)|}{|\sigma_1-\sigma_2|^\alpha} . \qquad (2.1)$$

При этом далее под L_α будем подразумевать лишь подпространство нечетных на $[-\pi,\pi]$ функций, чтобы удовлетворить условию действительности 2.

Обозначим

$$\Phi(z) = H(z) - \Lambda/z = D(z) + iM(z),$$

где $\Lambda = (\lambda_1,\dots,\lambda_N)$. Тогда формула обращения Гильберта /I7/ дает выражение действительной части $D(s)$ через мнимую $M(s)$ на контуре C_0

$$D(s) = \frac{1}{2\pi}\int_0^{2\pi} M(\varphi)\operatorname{ctg}\frac{\varphi-s}{2}d\varphi + D_0 \equiv KM + D_0 . \qquad (2.2)$$

Поэтому краевую задачу (I.3) можно рассматривать как операторное уравнение в пространстве

$$L_\alpha^N = \underbrace{L_\alpha \times \dots \times L_\alpha}_{N}, \quad \|x\|_{L_\alpha^N} = \max_j \|x_j\|_{L_\alpha}$$

относительно одной неизвестной вектор-функции $V(\varphi) = M(\varphi) - \Lambda \sin\varphi$

$$P(V) \equiv V - \bar{F}(\varphi) G(U, V) = 0, \quad \varphi \in [-\pi, \pi]. \qquad (2.3)$$

Здесь $U = K V + D_0$. При этом вектор D_0 можно задавать произвольно с одним условием $A D_0 = D_0$. Далее вектор D_0 будем считать фиксированным, также как и Λ .

Если $\bar{F}(\varphi) \in L_\alpha$, то оператор $P(V)$ определен в L_α^N, $P: L_\alpha^N \to L_\alpha^N$ и является дважды непрерывно дифференцируемым по Фреше. Первая производная есть ограниченный линейный оператор, представляющий собой оператор линейной задачи Римана-Гильберта для аналитического вектора $\Psi(z) = U(z) + i V(z)$ такого, что $U(0,0) = 0$:

$$P'(V_0) V = V(\varphi) - 2\bar{F}(\varphi)\left(\Omega(U_0) U(\varphi) + \Omega(V_0) V(\varphi)\right), \quad \varphi \in [0, 2\pi]. \qquad (2.4)$$

Матрица $\Omega(X)$ определяется по вектору $X(\varphi)$ выражением:

$$\Omega(X)_{ii} = x_i \; ; \quad \Omega_{ij} = 0, \; i \neq j \; ; \quad \varphi \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \qquad (2.5)$$

$$\Omega(X) = A T(X) A, \quad T_{ij} = 0, \; i \neq j \; ; \; T_{ii} = \sum_{j=1}^{N} A_{ij} x_j \; ; \; \varphi \in \left[\frac{\pi}{2}, \frac{3\pi}{2}\right].$$

Вторая производная представляет билинейную форму

$$P''(V_0)(V_1, V_2) = -2\bar{F}(\varphi) A\{T(U_1) A U_2 + T(V_1) A V_2\}. \qquad (2.6)$$

Для исследования уравнения (2.3) в /I4/ используется непрерывный аналог метода Ньютона

$$\frac{dV(\varphi, t)}{dt} = -P'(V)^{-1} P(V); \quad V(\varphi, 0) = V_0, \quad 0 \le t < \infty \qquad (2.7)$$

с помощью которого искомое решение получается как предел $V(\varphi, t)$ при $t \to \infty$. Условия сходимости этого процесса получены

в /I8/, а в более общем виде в /I9/.

В работах /5,6,I5/ исследован вопрос о существовании и единственности решений уравнения в форме (I.2) с помощью теоремы Шаудера и принципа сжатых отображений. Кроме того, в /5,6/ рассмотрены малые решения для N/D формулировки уравнения (I,2), в которой неизвестная функция $h_\alpha(w)$ ищется в виде $N(w)/D(w)$, а для функций N и D справедливы уравнения

$$D_\alpha(w)=1-\sum_{i=1}^{m(\alpha)}\frac{c_{\alpha i}}{w_{\alpha i}-w}-\frac{1}{\pi}\int_1^\infty\frac{\rho(w')N_\alpha(w')}{w'-w}dw';\quad w\in[1,\infty),\ t=1/w$$

$$N_\alpha(t)=B(t)-\sum_{i=1}^{m}c_i t_i t\,\frac{B(t)-B(t_i)}{t-t_i}-\frac{t}{\pi}\int_0^1\frac{B(t)-B(\tau)}{t-\tau}\frac{\rho(\tau)}{\tau}N(\tau)d\tau,\qquad(2.8)$$

$$B_\alpha(t)=\lambda_\alpha t+\frac{t}{\pi}\sum_\beta A_{\alpha\beta}\int_0^1\frac{\rho(\tau)|f_\beta(\tau)|^2}{\tau(\tau+t)}d\tau,$$

$$|h_\alpha(t)|^2=N^2(t)\left\{\left[1+\sum_i\frac{c_i t_i t}{t_i-t}+\frac{t}{\pi}\int_0^1\frac{\rho(\tau)N(\tau)}{\tau(\tau-t)}d\tau\right]^2+\rho^2(t)N^2(t)\right\}^{-1}.$$

Здесь $c_{\alpha i}$ - действительные числа, а точки t_i соответствуют положению КДД полюсов /I/. При этом знаки вычетов $c_{\alpha i}$ КДД полюсов подбираются так, что функции D_α не имеют нулей на разрезе $[0,1]$, т.е. D имеет представление

$$D(w)=R(w)E(z)=\prod_{i=1}^{m}\frac{1}{w-w_i}\exp\left[-\frac{z}{\pi}\int_1^\infty\frac{\delta(w')dw'}{w'(w'-w)}\right].$$

Основое достоинство N/D формулировки заключается в том, что задача сводится к системе уравнений типа Фредгольма. Кроме того, таким образом можно учесть КДД полюсы.

В /20/ для исследования "малых" решений уравнений $\pi-\pi$ рассеяния используется модифицированный метод Ньютона-Канторовича /2I/, однако, без обоснования сходимости.

Использование процесса (2.7) оправдывается тем, что область начальных приближений $V_0(\varphi)$, для которых траектории (2.7) стабилизируются к точному решению, вообще говоря, шире чем для дискретного метода Ньютона. Кроме того, в отличие от метода последовательных приближений, процесс (2.7) позволяет

построить решения с произвольным неотрицательным суммарным индексом производной Фреше.

Напомним, что суммарный индекс линейной краевой задачи Римана-Гильберта для аналитического вектора $H = U + iV$

$$B(\varphi)U(\varphi) - C(\varphi)V(\varphi) = D(\varphi), \quad \varphi \in [-\pi, \pi] \tag{2.9}$$

определяется согласно / 8/, как деленное на 2π приращение аргумента функции $\det[(B-iC)(B+iC)^{-1}]$ при обходе контура C_0 против часовой стрелки:

$$\varkappa = \frac{1}{2\pi}\Delta_0^{2\pi}\arg\{\det(B-iC)(B+iC)^{-1}\} = \frac{1}{\pi}\Delta_0^{2\pi}\arg\det(B-iC). \tag{2.10}$$

От величины $\varkappa$ зависит, будет ли краевая задача (2.9) безусловно и однозначно разрешима для всякой правой части $D(\varphi)$. В частности, для оператора $P'(V_0)$ из (2.4) можно получить $B - iC = iE - 2F(\varphi)\Omega(U_0 + iV_0)$; E - единичная матрица.

Теперь из формулы (2.13) для суммарного индекса легко вывести, что для всех решений $H_0(z) = U^0 + iV^0$ задачи I-5, удовлетворяющих условию

$$\max_{\varphi\in[0,\frac{\pi}{2}];\ i=1,\dots,N} |F(\varphi)V_i^0(\varphi)| < 1/2 \tag{2.11}$$

суммарный индекс производной Фреше равен нулю. Для этого рассмотрим, например, функцию $2F(\varphi)U_1^0 + i(2F(\varphi)V_1^0 - 1) \equiv b(\varphi)$. Так как $b(0) = -i$, а из условия 3 следует, что $|b(\varphi)| = 1$, то ввиду (2.II) кривая $b(\varphi)$ не пройдет через точку $z = i$, а значит, аргумент $b(\varphi)$ не получит приращения при обходе контура C_0.

Для расчета суммарного индекса, может быть использована следующая:

Лемма 2.I. Суммарный индекс $\varkappa$ линейной краевой задачи Римана-Гильберта

$$V - 2\bar{F}(\varphi)[\Omega(U_0)U + \Omega(V_0)V] = 0; \quad U_0, V_0 \in L_2^N \qquad (2.12)$$

вычисляется по формуле

$$\varkappa = \frac{2}{\pi}\sum_{j=1}^{N} \Delta_0^{\pi/2} \arg[2\bar{F}(\varphi)(U_j^\circ + iV_j^\circ) - i] \qquad (2.13)$$

если $U_0 = (U_1^\circ, \dots, U_N^\circ)$, $V_0 = (V_1^\circ, \dots, V_N^\circ)$ удовлетворяют условию 4.

Доказательство. В силу определения (2.10) имеем

$$\varkappa = \frac{1}{\pi}\Delta_0^{2\pi} \arg\{\det[2\bar{F}(\varphi)(\Omega(U_0) + i\Omega(V_0)) - iE]\}.$$

Согласно формулам (2.5) на правой полуокружности имеем:

$$\Delta_{-\frac{\pi}{2}}^{\pi/2} \arg\det(B - iC) = \sum_{j=1}^{N} \Delta_{-\frac{\pi}{2}}^{\pi/2} \arg[2\bar{F}(\varphi)(U_j^\circ + iV_j^\circ) - i] \qquad (2.14)$$

Используя свойство $A^2 = E$, приращение аргумента на $[\pi/2, 3\pi/2]$ можно вычислить по формуле

$$\Delta_{\frac{\pi}{2}}^{3\pi/2} \arg\det(B - iC) = \Delta_{\frac{\pi}{2}}^{3\pi/2} \arg\det A[2\bar{F}(T(U_0) + iT(V_0)) - iE]A =$$

$$= \Delta_{\pi/2}^{3\pi/2} \arg\prod_{j=1}^{N}\{2\bar{F}(\varphi)\sum_{K=1}^{N} A_{jK}(U_K^\circ + iV_K^\circ) - i\}.$$

Последнее выражение в силу условия 3 перекрестной симметрии принимает вид

$$\cdot \sum_{j=1}^{N} \Delta_{-\frac{\pi}{2}}^{\pi/2} \arg[2\bar{F}(\varphi)(U_j^\circ + iV_j^\circ) - i]. \qquad (2.15)$$

Складывая (2.14) и (2.15) и, учитывая симметрию относительно действительной оси, приходим к формуле (2.13). Лемма доказана.

Формула (2.13) показывает, что суммарный индекс $\varkappa$ зависит лишь от поведения решения $U_0 + iV_0$ на правой полуокружности. Это согласуется с тем фактом, что в подходе, связанном с интегральными уравнениями, вклад от левого разреза является вполне непрерывным оператором и не влияет на индекс /22/.

Отметим, что формула (2.13) в случае $N = 1$ позволяет легко установить область единственности решений уравнения (1.1).

Для этого достаточно, чтобы индекс линейной краевой задачи

$$V - 2\bar{F}(\varphi)(U_0 U + V_0 V) = 0$$

был равен нулю /I3/, что приводит к условию

$$\max_{\varphi\in[0,\pi/2]} |F(\varphi) V_0(\varphi)| < \frac{1}{2} .$$

В случае системы уравнений с $N > 1$ неизвестными функциями при таком подходе необходимо еще исследовать частные индексы системы /8/.

§ 3. Условия существования и единственности решений.

Как хорошо известно /I,4/, система (I.I) в случае одного и двух уравнений разрешима в замкнутом виде. Так в случае одного уравнения можно перейти к новой неизвестной функции

$$G(z) = h(z)^{-1},$$

тогда уравнение для $G(z)$ будет иметь вид

$$\operatorname{Im} G(\varphi) = -\bar{F}(\varphi), \quad 0 \le \varphi \le 2\pi,$$

откуда $G(z)$ находится при помощи интеграла Шварца.

Если число уравнений больше двух, то уже не удается построить общее решение системы (I.I). Некоторые классы решений для матриц A специального вида построены в /I0/. В большинстве случаев оказываются эффективными методы функционального анализа, позволяющие исследовать систему (I.I) для матриц A общего вида и при слабых ограничениях на функцию $F(\varphi)$.

Нелинейная краевая задача теории аналитических функций в случае одной неизвестной функции рассматривалась в /23,24/, где изучалось следующее уравнение

$$a(s)u(s) - \frac{b(s)}{2\pi}\int_0^{2\pi} u(\sigma)\operatorname{ctg}\frac{\sigma-s}{2}\,d\sigma = f_1(s) + \Phi_1\Big(s, u, -\int_0^{2\pi} u \operatorname{ctg}\frac{\sigma-s}{2}\,d\sigma, \lambda\Big), \quad (3.1)$$

которое сводилось к виду

$$U(s) = f(s) + \Phi\Big(s, U, -\frac{1}{2\pi}\int_0^{2\pi} U(\sigma)\operatorname{ctg}\frac{\sigma-s}{2}\,d\sigma + C, \lambda\Big), \quad (3.2)$$

удобному для применения принципа Шаудера и метода последовательных приближений.

Ограничения, налагаемые на функцию $\Phi(s,u,v,\lambda)$ (3.2) и гарантирующие существование решения, получены в /24/ для случая, когда $\varkappa = Ind\,[a(s)+ib(s)] = 0$. Случай отличного от нуля индекса $\varkappa$ исследован в /23/, где установлено существование решения, зависящего от $2\varkappa$ произвольных постоянных.

Непосредственное применение принципа Шаудера к системе уравнений в форме (I.2) проводится в /5,6/. Там же изучены уравнения в N/D формулировке (2.8) и в форме для обратных амплитуд. Полученные там оценки параметров $\Lambda = (\lambda_1, \ldots, \lambda_N)$, гарантирующих существование и единственность решения уравнения (I.2), для модели, рассмотренной в /3/, примерно в I0 раз меньше экспериментальных значений этих параметров.

Условия существования и единственности решений для уравнений (2.8), полученные с помощью метода Ньютона-Канторовича /25/, оказываются близкими к условиям из /5-6/.

Остановимся кратко на непрерывном процессе Ньютона (2.7), применявшемся к задаче I-5 в /I4/.

Введем следующие обозначения:

$$\|\bar{F}(\varphi)\|_{L_\alpha} = F_\alpha \;; \quad a = \|A\| = \max_{1\le i\le N} \left(\sum_{j=1}^{N} |A_{ij}|\right).$$

Пусть a^* есть максимальное собственное значение матрицы AA^*. Если $A = A^*$, как предполагалось в /6/, то $a^* = I$. Обозначим норму сингулярного оператора K из (2.2) через K_α, $\|K\| = K_\alpha$. Согласно оценке из /26/ для нормы сингулярного интеграла типа Коши нетрудно получить неравенство

$$K_\alpha \le 2\left(2 + \frac{2^{1+\alpha}}{\pi\alpha} + \frac{2}{\pi(1-\alpha)} + \left(\frac{2\pi}{\alpha}\right)^{\alpha-1}\right).$$

Определим на окружности C_0 вектор-функцию $R(\varphi, \Lambda, D_0)$

$$R(\varphi,\Lambda,D_0)\equiv P\left(\frac{\Lambda}{2}+D_0\right)=\begin{cases}-\Lambda\sin\varphi-\bar F(\varphi)\left[\Lambda^2+D_0^2+2\Lambda D_0\cos\varphi\right], & \varphi\in\left[-\frac{\pi}{2},\frac{\pi}{2}\right]\\ -\Lambda\sin\varphi-\bar F(\varphi)\left[A(\Lambda^2+D_0^2)+2\Lambda D_0\cos\varphi\right], & \varphi\in\left[\frac{\pi}{2},\frac{3\pi}{2}\right]\end{cases}$$

Справедлива следующая

Теорема 3.I. а) пусть векторы $\Lambda_0, D_0\in R^N$ таковы, что $AD_0=D_0$, $A\Lambda_0=-\Lambda_0$ и выполнено соотношение

$$\|R(\varphi,\Lambda_0,D_0)\|_{L_\alpha^N}\le\left[8F_\alpha a^3(1+K_\alpha^2)\right]^{-1}. \qquad (3.3)$$

Тогда существует решение $\Phi(z)=H(z)-\frac{\Lambda_0}{2}$ задачи I-5 при $\Lambda=\Lambda_0$ и $\Phi(0,0)=D_0$ такое, что $\|Im\,\Phi\|_{L_\alpha^N}\le\left[4F_\alpha a^3(1+K_\alpha^2)\right]^{-1}$. Это решение можно получить с помощью процесса (2.7) с начальным приближением $V_0(\varphi)=0$.

б) при $\Lambda=0$, $D_0=0$ в области

$$\Omega=\left\{V(\varphi)\in L_\alpha^N \Big/ \max_{\varphi\in[0,\frac{\pi}{2}],\ i\le N}|F(\varphi)V_i(\varphi)|<\frac{1}{2a^*}\right\} \qquad (3.4)$$

нет других решений, кроме $H(z)=0$. Решение для $\Lambda\neq0$ единственно в области

$$\Omega_1=\left\{V(\varphi)\in L_\alpha^N \Big/ \max_{\varphi\in[0,\frac{\pi}{2}],\ i\le N}|F(\varphi)V_i(\varphi)|<\frac{1}{4a^*}\right\}. \qquad (3.5)$$

Доказательство первой части теоремы сводится к проверке условий, обеспечивающих сходимость процесса (2.7) /I8/.

Для обоснования условий единственности отметим, что если существуют два решения U_1+iV_1 и U_2+iV_2 с одинаковыми Λ_0, D_0, то их разность $\Delta H=\Delta U+i\Delta V$ является аналитической, исчезающей в нуле функцией, и для нее справедливо представление

$$\Delta V-\bar F(\varphi)\left[AT(U_1+U_2)AK\Delta V+AT(V_1+V_2)A\Delta V\right]\equiv\Psi(\Delta V)=0. \qquad (3.6)$$

Рассматривая уравнение (3.6) в пространстве $L_2^N(C_0)$, и, учитывая, что $\|K\|_{L_2^N(C_0)}=I$ /26/, из (3.4), (3.5) получается оценка

$\|\Psi\|_{[L_2^N \to L_2^N]} < 1$, что и доказывает единственность.

Отметим, что для всякого решения задачи I-5 выполнено неравенство

$$\max_{\varphi\in[0,\frac{\pi}{2}],\, i\le N} |F(\varphi)\, V_i(\varphi)| \le 1$$

которое показывает, что области единственности (3.4),(3.5) являются существенно нелокальными.

В работе /I5/ из уравнения (I.2) получено уравнение для $V_i(t) = \frac{Im\, h_i(t)}{t}$:

$$V = t\rho(t)\left[V^2 + \left(\lambda - \frac{1}{\pi}\int_0^1 \frac{V(\tau)}{\tau - t}d\tau + A\frac{1}{\pi}\int_0^1 \frac{V(\tau)}{\tau + t}d\tau\right)^2\right], \quad (3.7)$$

при анализе которого установлены менее ограничительные по сравнению с /5-6/ условия существования и единственности решений (I.2). Использование теоремы Шаудера при этом существенно опирается на положительность решения $V(t) \ge 0$. В результате удалось ослабить требование как на функцию $\rho(t)$, так и на матрицу A.

Определим множество G следующим образом

$$G = \{V \in L_\alpha^N[0,1] \,/\, V_i(t) \ge 0;\ V_i(0) = V_i(1) = 0;\ \lim_{t\to 0} t^{-2} V(t) < \infty\}.$$

Используем следующие константы

$$N_\alpha = \frac{2}{(1-\alpha)\pi} + \frac{1 + 2^{1+\alpha} + 3^\alpha}{\pi\alpha};\quad \|A^+\| = \max_{i\le N}\left(\sum_{j=1,\, A_{ij}>0}^{N} A_{ij}\right);$$

$$\|\rho\| = \frac{1}{\pi}\|\rho(t)\|_{L_2};\quad b = \|t\rho(t)\|_{C[0,1]};\quad d = \|t\rho(t)\|_{L_\alpha};\quad |\Lambda| = \max_i |\lambda_i|,$$

величины α и α^* имеют прежний смысл.

<u>Теорема 3.2.</u> а) Если число $R > 0$ таково, что

$$2b\left\{\left[|\Lambda| + (\alpha+1)\frac{R}{\pi\alpha}\right]\left(\frac{\alpha}{\pi} + N_\alpha\right)R + R^2\right\} + \left\{\left[|\Lambda| + (\alpha+1)\frac{R}{\pi\alpha}\right]^2 + R^2\right\}d \le R, \quad (3.8)$$

то на отрезке $\|V\|_{L_\alpha^N} \le R$ конуса G уравнение (3.7) имеет хотя бы одно решение.

б) Пусть число $R_1 > 0$ таково, что

$$\left(|\Lambda| + (a+1)\frac{R_1}{\pi\alpha}\right)\left(a^*\|p\| + b\right) + R_1 b < \frac{1}{2}, \qquad (3.9)$$

тогда на отрезке $\|V\|_{L_\infty^N} < R_1$ конуса G уравнение (3.7) имеет не более одного решения.

Доказательство этой теоремы можно найти в /I5/. Там же показано, что оценки (3.8), (3,9) дают более широкие области существования и единственности, нежели условия из /5-6/.

В табл.I приводятся различные оценки для константы связи $|\Lambda|$ при следующих данных

$$\rho(t) = \frac{\kappa^3}{12\pi}\exp\left(-\frac{\kappa^2}{49}\right), \quad \kappa = \left(\frac{1}{t^2} - 1\right)^{1/2}, \quad 0 \le t \le 1,$$

$$A = \frac{1}{9}\begin{pmatrix} 1 & -8 & 16 \\ -2 & 7 & 4 \\ 4 & 4 & 1 \end{pmatrix}, \qquad (3.10)$$

которые рассматривались в ряде работ /5,6,27,3/. Отметим, что если матрица A удовлетворяет условию

$$\sum_{\beta=1}^{N} A_{\alpha\beta} = 1; \quad \alpha = 1, \ldots, N, \qquad (3.11)$$

то вместо величины a в (3.8), (3.9) можно использовать $\|A^+\| < a$, что улучшает эти оценки для $|\Lambda|$ и R.

Решения, гарантируемые предыдущими теоремами, имеют, очевидно, нулевой суммарный индекс (п.2) производной Фреше. Установим далее существование решений уравнения (I.I), имеющих произвольный неотрицательный суммарный индекс $\varkappa$ производной Фреше. Эти решения лежат вне области Ω_1 (3.5) и не могут быть получены с помощью теоремы о сжатом отображении. Рассмотрим матрицы A, удовлетяорющие условию (3.II). Тогда, если $h(z)$ является решением задачи I-5 при N =I, A = I, то вектор $H_0(z) = (h(z), \ldots, h(z))$, имеющий одинаковые компоненты, будет решением этой задачи для произвольного $N >$ I. Назовем

это решение диагональным.

Установим связь между величиной $\varkappa$ для диагонального решения и полюсами функции $h(z)$ в плоскости z . Обозначим через $P_\psi(\Omega)$ число полюсов функции ψ в области Ω . Рассмотрим следующие области: $\Omega_1=\{z:|z|>1\}$; $\Omega_2=\{z:|z|<1\}$; $\Omega_3=\{z:|z|>1, Re\,z=0\}$; $\Omega_4=\{z:|z|>1, Re\,z>0\}$; $\Omega_5=\{z:|z|<1, Re\,z=0\}$; $\Omega_6=\{z:|z|<1, Re\,z>0\}$. Определим целые числа κ, p_0, p_1, l_0, l_1 формулами $l_1=P_h(\Omega_6)$, $\kappa=P_F(\Omega_1)-P_F(\Omega_2)$; $p_0=P_h(\Omega_3)$; $p_1=P_h(\Omega_4)$; $l_0=P_h(\Omega_5)$.

Имеет место

Лемма 3.I. Суммарный индекс $\varkappa$ производной Фреше (2.4) на диагональном решении выражается формулой

$$\varkappa=N(p_0+2p_1-l_0-2l_1+\kappa). \qquad (3.I2)$$

Доказательство леммы можно получить, если рассмотреть приращение аргумента функции $S(z)=(s_1,\dots,s_N)$,

$$S_j(z)=1+2i\,F^*(z)h_j(z), \quad j=1,\dots,N$$

на контуре, охватывающем правую полуокружность, и обходящем полюса и нули $S(z)$ на действительной оси /I4/. Такой контур рассматривался в /28/. При этом функция $\bar{F}(\varphi)$ есть краевое значение на Со мероморфной функции $F(z)$. Кроме того используется аналитическое продолжение $S(z)$ на внешность единичного круга Со

$$S(z)=\left[S\left(\tfrac{1}{z}\right)\right]^{-1}, \quad |z|>1, \quad Re\,z\geq 0$$

и формула (2.I2), переписанная в виде

$$\varkappa=\frac{4}{\pi}\sum_{j=1}^{N}(\delta_j(i)-\delta_j(0)); \quad S_j(\sigma)=\exp[2i\,\delta_j(\sigma)], \ \sigma=e^{i\varphi}, \ \varphi\in[0,\tfrac{\pi}{2}].$$

В /I4/ формула (3.I2) обобщается для произвольных решений с двухрядной матрицей А вида

$$A=\begin{pmatrix} a & 1-a \\ 1+a & -a \end{pmatrix}, \quad |a|\leq 1.$$

В случае отличного от нуля суммарного индекса производная Фреше не имеет ограниченного обратного оператора, а решение зависит от конечного числа произвольных параметров. Если зафиксировать согласно формуле (3.12) определенное число полюсов функции $H(z)$ вне круга Co, то можно получить единственное решение.

Теорема 3.3. Пусть векторы Λ_0, D_0 таковы, что число $\mathcal{E} = \|G_N(H_0, \Lambda_0, D_0)\|$, определенное для диагонального решения $H_0 = U_0 + i V_0$, удовлетворяет условию $2\mathcal{E}M^2L < 1$. Тогда для матриц A типа (3.11) и произвольного $m > 0$ задача I-5 имеет в шаре $\|V - V_0\| \le (1 - \sqrt{2\mathcal{E}M^2L})(ML)^{-1}$ единственное решение, суммарный индекс производной Фреше для которого равен m, если фиксировать m полюсов функции $H(z) = U + iV$ (либо $m/2$ полюсов вместе с их вычетами) вне круга Co. Это решение можно получить методом Ньютона с начальным приближением $H_0(z)$.

Здесь $H_0(z)$ - известное точное решение задачи I-5 с равными компонентами, G_N -нелинейный оператор, соответствующий задаче I-5 вместе с указанными дополнительными условиями; $L = \|G_N''(H)\|$; $M \ge \|G_N'(H_0)^{-1}\|$, а $D_0 = H(z) - \frac{\Lambda_0}{z}\Big|_{z=0}$.

Доказательство теоремы использует явный вид однокомпонентного решения, а также формулу для точного решения одномерной линейной задачи Гильберта /17/. Подробное доказательство приводится в /14/.

Отметим, что малые по норме решения (теорема 3.1) получаются здесь как частный случай. Кроме того теорема 3.3. непосредственно указывает и метод расчета больших по модулю решений.

§ 4. Приближенное решение уравнений типа Лоу.

Численные расчеты уравнений Лоу можно проводить как для уравнений в форме (1.2), так и для краевой задачи I-5. Используя хорошо известный метод моментов /17/, краевую задачу I-5

можно заменить бесконечной алгебраической системой для коэффициентов ряда Лорана функции $H(z)$

$$h_\alpha(z) = \sum_{n=-\infty}^{\infty} a_n^\alpha z^n \quad , \quad \alpha = 1, \dots, N \, .$$

Для уравнения Лоу такой подход впервые использовался в /7,29/. Согласно /7/ задача I-5 эквивалентна системе нелинейных алгебраических уравнений:

$$a_\nu^\alpha = a_{-\nu}^\alpha + \sum_{k=-\infty}^{\infty} F(\nu,k) \sum_{m=-\infty}^{\infty} E_\nu^\alpha (a_m^\alpha ; a_{m+k}^\alpha) ;$$

$$\alpha = 1, \dots, N \; ; \quad \nu = 1, 2, \dots \infty \, , \qquad (4.1)$$

где

$$F(\nu,k) = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \sin \nu\varphi \cos k\varphi \, F(\varphi) \, d\varphi ,$$

$$E_\nu^\alpha (a_m ; a_{m+k}) = a_m^\alpha a_{m+k}^\alpha + (-1)^\nu \sum_{\beta=1}^{N} A_{\alpha\beta} a_m^\beta a_{m+k}^\beta \, . \qquad (4.2)$$

Кроме того $a_{-1}^\alpha = \lambda_\alpha$, $a_{-\nu}^\alpha = 0$, $\nu \geq 2$.Вектор $D_0 = (a_0^\nu)$, $\alpha = 1, \dots, N$ считается заданным, так что $AD_0 = D_0$.

Предполагая, что краевое значение $H(b)$, $b \in C_0$, удовлетворяет условию Гельдера (2.I) $H(b) \in L_\mu^N$, можно заключить, согласно /7,30/ что коэффициенты a_ν^α удовлетворяют ограничениям

$$a_\nu^\alpha = O\left(\nu^{-(\mu+1)}\right) \, , \quad \alpha = 1, \dots, N. \qquad (4.3)$$

Таким образом система (4.I) определяет в пространстве ℓ_μ бесконечных последовательностей $Y = (y_1, \dots, y_n, \dots)$, $y_n \in R^N$, удовлетворяющих условию

$$|y_n| \leq C n^{-(1+\mu)} \, , \quad \|Y\|_{\ell_\mu} = \sup_n |y_n| \qquad (4.4)$$

операторное уравнение

$$Y = \Lambda + A(Y) \, , \qquad (4.5)$$

где оператор A вполне непрерывен в ℓ_μ /7/. Разрешимость уравнения (4.5) при малых $\Lambda = (\lambda_\alpha)$ доказана в /7/, а также

следует из теоремы 3.I. Если положить $a_{\nu}^{\alpha}=0$, $\nu>\kappa$, то система (4.I) превратится в конечномерную и будет также однозначно разрешима при малых Λ . Пусть вектор Y_κ получается из Y , если положить $y_n=0$, $n>\kappa$. Обозначив решение конечной системы через $[Y]_\kappa$, а соответствующий А конечномерный оператор через $[A]_\kappa$, нетрудно видеть, что

$$[Y]_\kappa - Y_\kappa = [A]_\kappa [Y]_\kappa - [A]_\kappa Y_\kappa + \kappa^{-\mu} B\,; \quad B=O(1),$$

откуда получается оценка погрешности приближенного решения

$$\| Y_\kappa - [Y]_\kappa \| \le C_1 \kappa^{-\mu}, \quad \|x\| = \sup_{n\le\kappa} |x_n|. \qquad (4.6)$$

В работах /7,29/ указанная алгебраическая система решалась методом простой итерации. Более предпочтительными здесь являются итерационные процессы типа Ньютона, позволяющие рассчитывать решения с произвольным суммарным индексом производной Фреше. В /I4/ наряду с процессом (2.7) использовался метод наискорейшего спуска

$$\frac{dV(\varphi,t)}{dt} = -[P'(V)]^* P(V)\,; \quad V(\varphi,0)=V_0\,, \quad 0\le t<\infty \qquad (4.7)$$

с помощью которого для матрицы

$$A=\begin{pmatrix}0&1\\1&0\end{pmatrix} \qquad (4.8)$$

и функции $F(\varphi)=\frac{1}{2}\sin 2\varphi$ было рассчитано решение, имеющее суммарный индекс, равный 8:

$$h^{\alpha}(z)^{-1} = \frac{2z}{\lambda_\alpha(1+z^2-\varepsilon_0^\alpha z)} - \frac{z^2}{2} - \frac{4z^2R}{(1+z^2)^2\omega^2-4z^2}\,. \qquad (4.8')$$

Это решение получается "возмущением" диагонального

$$h(z)^{-1} = C - \frac{z^2}{2} - \frac{4Rz^2}{(1+z^2)\omega^2-4z^2} \qquad (4.9)$$

при следующем соотношении параметров: $\varepsilon_0^\alpha = -\frac{2}{C\lambda_\alpha}$.

Если $\lambda_1 = 0.01$; ε_0^{α} =-173,9 ; $\mathcal{R}$ =-2, ω^2 =9, то функция (4.8) имеет полюс при z =1,137, с вычетом -.9856. При дополнительных условиях, указанных в теореме 3,3. и при начальном условии (4.9) процесс (4.7) сходится к решению (хотя и очень медленно). Точность полученных решений проиллюстрирована в табл.2, где представлены точные и приближенные значения a_ν' для $\nu \le 70$.

Непосредственное использование уравнений (I.2) (в форме (3.7)) предполагает построение квадратурных формул для сингулярного интеграла типа Коши. Как показано в /I6/, гладкость решения $h_\alpha(t)$ определяется гладкостью функции $g(t) = t\rho(t)$, $t \in [0,1]$. Пусть $H^{\alpha+m}$ (m - целое число, $0 < \alpha < 1$) - B - пространство вектор функций, непрерывных по Гельдеру вместе со своими производными до m -ого порядка включительно, с нормой

$$\|x\| = \max_{i \le N} \left\{ \max_{0 \le t \le 1} |x_i^{(m)}(t)| + \sup \frac{|x_i^{(m)}(t) - x_i^{(m)}(t_1)|}{|t - t_1|^{\alpha}} \right\}$$

и таких, что $x(0) = x(1) = 0$. Пусть $g(t) \in H^{\alpha+m}$. Точность приближенных решений оценивается в /I6/ для следующих квадратурных формул:

а) $m = 0$,

$$\left\{ \frac{1}{\pi} \int_0^1 \frac{t\,x(t)}{t - t_j}\,dt \right\}_{j=1}^{n} \approx \left\{ \frac{1}{\pi} t_j x_{ij} \ln \frac{1 - t_j}{t_j} + \right.$$

$$\left. + \frac{1}{\pi} \sum_{\substack{\kappa=1 \\ \kappa \ne j, j-1}}^{n} (t_{\kappa+1} - t_\kappa) \frac{t_\kappa x_{i\kappa} - t_j x_{ij}}{t_\kappa - t_j} \right\}_{j=1,n}^{i=1,N} . \qquad (4.10)$$

Узлы t_j выбираются по формуле $t_j = \frac{1}{2} + \frac{1}{2} d_j$, где d_j - точки отрезка $[-1, 1]$, расположенные симметрично относительно точки $t = 0$ /3I/.

б) $m \ge 1$, тогда используется формула, предложенная в /32/:

$$\left\{ \frac{1}{\pi} \int_0^1 \frac{t\,x(t)}{t - t_j}\,dt \right\}_{j=1,n} \approx \left\{ -\frac{1}{\pi} \sum_{\kappa=1}^{n} \frac{t_\kappa x_{ij}}{P_n'(t_\kappa)} \cdot \frac{Q_n(t_\kappa) - Q_n(t_j)}{t_\kappa - t_j} \right\}_{j=1,n}^{i=1,N} , \qquad (4.11)$$

где при $\kappa = j$ вместо

$$\frac{q_n(t_\kappa) - q_n(t_j)}{t_\kappa - t_j}$$

надо взять $q_n'(t_\kappa)$. Здесь $P_e(t)$ - полиномы Лежандра:

$$P_e(t) = \left(\frac{2e+1}{2}\right)^{1/2} \frac{1}{2^e e!} \frac{d^e (t^2-1)^e}{dt^e}; \quad q_e(t) = -\int_{-1}^{1} \frac{P_e(t')\,dt'}{t'-t}.$$

Для вычисления регулярного интеграла в обоих случаях в /16/ используется квадратурная формула Гаусса /33/, а для построения приближенных решений уравнения (3.7) используется параболический сплайн /34/. В итоге имеет место

Теорема 4.1. Пусть $g(t) \in H^{\alpha+m}$, а величина $|\lambda|$ достаточно мала. Тогда в некотором шаре $\|x(t)\| \le R$, $R > 0$ пространства $H^{\alpha+m}$ уравнение (3.7) имеет единственное решение $x^*(t)$, которое может быть получено как предел последовательности $x^{(n)} = \bar{\varphi}_n x_n^*$, где $\bar{\varphi}_n$ - указанные сплайны, а x_n^* - единственное решение (при всяком n) алгебраической системы уравнений

$$x_n = A x_n \equiv g(t) / [x_n^2 + \bar{U}^2(x_n)],$$

получающейся из (3.7) с помощью квадратурных формул (4.10), (4.11)

При этом: 1) Процесс нахождения каркасов /35/ x_n^* приближенных решений сходится и справедлива оценка

$$\|\varphi_n x^* - x_n^*\| \le \frac{1}{1-q} \begin{cases} M_1 n^{-\alpha} \ln 2n + \|A\| M_2 (2n+1)^{-\alpha}, & m = 0 \\ M_4 (2n-1)^{-m+1-\alpha} + \|A\| M_5 (2n+1)^{-m-\alpha}, & m \ge 1. \end{cases} \quad (4.12)$$

2) Процесс нахождения приближенных решений сходится со скоростью

$$\|x^{(n)} - x^*\| \le \begin{cases} M_3 \|\Delta_n\|^\alpha + \|\bar{\varphi}_n\| \cdot \|\varphi_n x^* - x_n^*\|, & m = 0 \\ M_6 n^{-m} + \|\bar{\varphi}_n\| \, \|\varphi_n x^* - x_n^*\|, & m \ge 1, \end{cases} \quad (4.13)$$

где $\|\Delta_n\| = \max_j |t_{j+1} - t_j|$.

Одной из задач, которые рассчитывались с помощью уравнений (3.7) была система с трехрядной матрицей (3.10) и указанной там же функцией $\rho(t)$. Начиная со значения $|\lambda|$, гарантирующе-

го существование и единственность решения, решаем систему (3.7) методом наискорейшего спуска. При переходе к большей величине параметра Λ используем начальное приближение, совпадающее с уже найденным решением для предыдущего значения. Такой процесс удалось продолжить до значения $|\Lambda| = 7{,}8$, в то время как соответствующее экспериментальное значение константы 8,7 /3/.

О точности приближенных решений можно судить по расчетам для системы (4.8), однозначное решение для которой существует лишь при $|\Lambda| < 2$. Сравнение точного решения

$$V_\alpha(t) = 4t^2\lambda_\alpha^2\sqrt{1-t^2}\left(4 - 4\lambda_\alpha t^2(2t^2-1) + \lambda_\alpha^2 t^2\right)^{-1}, \quad \alpha = 1,2 \qquad (4.15)$$

и приближенного решения (3.7) приведено в табл.3, в которой можно наблюдать, как падает точность расчетов при увеличении гельдеровской нормы решения $h_\alpha(t)$, при $\mu = 1/2$.

В заключение хотелось бы отметить, что опыт, накопленный в практике численного решения уравнений типа Чу-Лоу, показывает, что наибольшие трудности возникают при расчете решений с отличным от нуля суммарным индексом производной Фреше, причем эти трудности сохраняются и в N/D формулировке. Расчет малых решений путем постепенного увеличения параметра Λ обычно позволяет найти все семейство решений с малой нормой вплоть до значения Λ, близкого к Λ_{max}. При этом, если $\Lambda \to \Lambda_{max}$, то точность приближенных решений значительно ухудшается в связи с ростом констант M_1, M_2, M_3, M_4 в оценках (4.12), (4.13). Для уточнения решений здесь можно использовать интерполяцию по Ричардсону.

Таблица I.

Оценки сверху константы связи /I/.

	Работа /I5/	Работы /5,6/		
		Уравнение в форме (I.2)	Уравнение для обратных амплитуд	N/D метод
Область существования решений	0,I0	0,0I4	0 0I4	0,II
Область единственности решений	0,20	0,004I	0,006I	0,05I

Таблица 2.

Коэффициенты a_ν^1 с четными номерами при λ =0,0I ; ε^1 =-I73,9; R =-2,0; ω^2 =9,0; $\varkappa$ =70

ν	a_ν^1 для(4,8')	ν	a_ν^1 для (4,I)	ν	a_ν^1для(4,8')	ν	a_ν^1для (4.I)
0	.86957E+00	0	.8697IE+00	36	.93944E-02	36	.968I5E-02
2	.I7223E+0I	2	.I724IE+0I	38	,I3340E-0I	38	.I3568E-0I
4	.I3203E+0I	4	.I3229E+0I	40	.59992E-02	40	.6I832E-02
6	.38I89E+00	6	.38482E+00	42	,23I72E-02	42	.24654E-02
8	.48384E+00	8	.48678E+00	44	.48432E-02	44	.49604E-02
I0	.656I3E+00	I0	.65890E+00	46	.39I73E-02	46	.40I02E-02
I2	.28837E+00	I2	.2909IE+00	48	.II052E-02	48	.II797E-02
I4	.I2059E+00	I4	.I2285E+00	50	.I3077E-02	50	.I3668E-02
I6	.24320E+00	I6	.245I5E+00	52	.I90I4E-02	52	.I9477E-02
I8	.I8974E+00	I8	.I9I40E+00	54	.86557E-03	54	.90236E-03
20	.54373E-0I	20	.55792E-0I	56	.32III E-03	56	.35036E-03
22	.67442E-0I	22	.68628E-0I	58	.68303E-03	58	.7059IE-03
24	.93567E-0I	24	.94542E-0I	60	.56277E-03	60	.5807IE-03
26	.4I588E-0I	26	.42397E-0I	62	.I5778E-03	62	.I7203E-03
28	.I67I8E-0I	28	.I7385E-0I	64	.I8I89E-03	64	.I9307E-03
30	.34327E-0I	30	.34867E-0I	66	.27094E-03	66	.27960E-03
32	.27264E-0I	32	.2770IE-0I	68	.I249IE-03	68	.I3236E-03
34	.77485E-02	34	.8I050E-02				

Таблица 3.

Точные и приближенные решения $V_i(t)$ для /4,I5/.

t	λ =0,05		λ =0,65		λ =I,35	
	точное	приближенное	точное	приближенное	точное	приближен
0,I	.2466E-04	.2467E-64	.4I73E-02	.7472E-02	.I78E-0I	.205E-0I
0.2	.9779E-04	.9798E-04	.I8IIE-0I	.2072E-0I	.669E-0I	.750E-0I
0.3	.2I38E-03	.2I46E-03	.3430E-0I	.3879E-0I	.I37E-00	.I64E+00
0.4	.3646E-03	.3666E-03	.5697E-0I	.6254E-0I	.2I9E-00	.240E+00
0.5	.5778E-03	.54I3E-03	.8258E-0I	.8949E-0I	.308E+00	.34IE+00
0.6	.7I62E-03	.7200E-03	.II03E+00	.I202E+00	.404E+00	.440E+00
0.7	.874IE-03	.8748E-03	.I397E+00	.I538E+00	.5I63+00	.572E+00
0.8	.9662E-03	.9600E-03	.I706E+00	.I87IE+00	.667E+00	.725E+00
0,9	.9049E-03	.8627E-03	.I965E+00	.2048E+00	.93IE+00	.I02E+0I

ЦИТИРОВАННАЯ ЛИТЕРАТУРА:

I. L.Castillejo,R.Dalitz,F.Dyson.Low's scattering equation for the charged and neutral scalar theories. Phys.Rev.,1956,101, N°1,453-458.

2. G.Shem,S.Mandelstam.Theory of the Low-energy pion-pion interaction.Phys.Rev.,1960,119,467-477.

3. G.Saltzman,F.Saltzman.Solution of the static theory integral forpion-nucleon scattering in the one-meson approximation Phys.Rev.,1957,108,1619-1628.

4. В.И.Журавлев, В.А.Мещеряков. Статические модели в дисперсионном подходе. ЭЧАЯ, I974, 5, вып.I, I72-222.

5. R.L.Warnock.Existence proof by a fixed-point theorem for solution of the Low equation.Phys.Rev.,1968,170,1323-1331; Phys.Rev.,1968,174,2169-2169.

6. H.Mc Daniel,R.L.Warnock.Bounds on coupling constants in existence theorems for the Low equation.Nuovo Cimento,1969, 64,905-926.

7. I.P.Nedelkov.Low's problem and its investigation by means of power-series expansion.Phys.Rev.,1972.D6,2842-2852. I.P.Nedelkov.J.Math.Phys.,Vol.15,N°9,1974.

8. Н.П.Векуа. Системы сингулярных интегральных уравнений. М., "Наука", I966.

9. Н.И. Мусхелишвили. Сингулярные интегральные уравнения.М., "Наука", I966.

I0. В.А.Мещеряков. Метод построения некоторых классов решений уравнений типа уравнений Чу-Лоу. ОИЯИ, Р-2369, Дубна, I965.

II. В.П.Гердт, В.А.Мещеряков. Локальный вид решений уравнения Чу-Лоу. Теоретич.и матем. физ., 24, № 2, I55-I63.

I2. В.П.Гердт. Аналитическое вычисление инвариантной кривой уравнения Чу-Лоу ОИЯИ, Р4-I2064, Дубна, I978.

I3. Е.П.Жидков, И.П.Недялков, Б.Н.Хоромский. Об одной нелинейной краевой задаче для аналитических функций, связанной с уравнениями типа Чу-Лоу. Тр.Междунар.Совещания по программированию и матем.методам решения физ.задач. Дубна, ОИЯИ, ДI0-II264, I978.

I4. Е.П.Жидков, М.Нгуен, И.П.Недялков, Б.Н.Хоромский. Исследование одного класса решений уравнения Лоу, ч.I. Малые по модулю решения. ОИЯИ, Р5-II470, Дубна, I978; ч.II. Большие по модулю решения. ОИЯИ, Р5-II47I, Дубна, I978; ЖВМ и МФ, I979, № 4.

I5. Е.П.Жидков, М.Нгуен, Б.Н.Хоромский. Качественное исследование и приближенное решение нерегуляризованного уравнения Лоу. ОИЯИ, Р5-II9I2, Дубна, I978.

I6. Е.П.Жидков, М.Нгуен, Б.Н.Хоромский. О сходимости итерационных процессов приближенного решения нелинейного сингулярного интегрального уравнения Лоу. ОИЯИ, Р11-I2247, Дубна, I979.

I7. Ф.Д.Гахов. Краевые задачи. М., "Наука", I977.

I8. М.К.Гавурин. Нелинейные функциональные уравнения и непрерывные аналоги итеративных процессов. Изв. вузов.Математика, № 5, I958, I8-36.

I9. Б.Н.Хоромский. Автореферат канд.диссертации, ОИЯИ, 5-II643, Дубна, I978.

20. A.T.Amatuni.A possible application of Newton-Kantorovich's method to Low-energy $\pi-\pi$ -scattering.Nuovo Cimento, 1969, 58A,321-339.

2I. Л.В.Канторович, Г.П.Акилов. Функциональный анализ.М.,"Наука", I977.

22. C.Lovelace.Uniqueness and Symmetry breaking in S-matrix theory. Communs.Math.Phys.,1967,4,N°4,261-302.

23. В.К.Наталевич. О нелинейном сингулярном интегральном уравнении и нелинейной краевой задаче теории аналитических функций. ДАН СССР, т.83, № I, I952, с.I9-22.

24. А.И.Гусейнов, Матем.сборн., 26 (68):2 (I950)

25. Конно Тайра, Н.Ф.Нелипа. Применение метода Ньютона-Канторовича к N/D уравнениям Чу-Лоу. I. Теоретич. и матем.физ., I2, № 2, 2I4-222.

26. И.И.Данилюк. Нерегулярные граничные задачи на плоскости. М., "Наука", I975.

27. И.П.Недялков, Г.И.Пенчев. ОИЯИ, Р-I445, Дубна, I963.

28. K.Huang,A.H.Mueller.Exact bootstrap solution to the Low equation.Phys.Rev.,1965,140B,365-374.

29. I.P.Nedelkov.A model equation for non-analytical transition amplitudes.Dubna,1963,JINR,E-1294.

Н.Х.Бырнев, В.А.Мещеряков, И.П.Недялков. Об одной алгебраической системе, эквивалентной уравнениям Лоу. Ж.эксперим. и теор. физ. I964, 46, 663-670.

30. Н.К.Бари. Тригонометрические ряды. М., ФИЗМАТГИЗ, I96I.

3I. В.В.Иванов. Теория приближенных методов и ее применение к численному решению сингулярных интегральных уравнений, Киев, "Наукова Думка", I968.

32. А.А.Корнейчук. ОИЯИ, Р-I3I7, Дубна, I963.

33. Н.С.Бахвалов. Численные методы. М., "Наука", I975.

34. С.Б.Стечкин, Ю.Н.Субботин. Сплайны в вычислительной математике, М., "Наука", I976

35. А.А.Самарский. Теория разностных схем. М., "Наука", I977.

NOTES ON GENERALIZED PADÉ APPROXIMATION

by

G. Németh

Central Research Institute for Physics,

Budapest

АННОТАЦИЯ

В работе представлены примеры решения обобщенной аппроксимации Паде в явном виде для некоторых специальных функций. Исходя из известных разложений функций arctg x, ln/1+x/ и $x^{1/2}$ в конечном отрезке, можно определить хорошо аппроксимирующие рациональные дроби.Для функции e^{-x} на бесконечном интервале /o,∞/ также получены рациональные дроби с помощью обобщенного метода Паде.

CONTENT

In this paper explicit solutions are presented to show the power of the generalized Padé approximations for some special functions. For functions arctg x and ln/1+x/ in finite interval excellent rational approximations are determined. Also, for e^{-x} in /o,∞/ a quite good rational approximation is obtained.

1.In the last few years the theory of rational approximation has shown great development / see e.g. Reddy[1]/.One of the most powerful methods of obtaining rational approximations is Padé's method which is as follows. Let

$$f(x) \sim c_0 + c_1 x + c_2 x^2 + \dots , \quad x \to 0,$$

be a formal power series. It is possible to determine polynomials $p_n(x)$ and $q_m(x)$ such that

$$f(x) - \frac{p_n(x)}{q_m(x)} = O(x^{n+m+1}) , \quad x \to 0, \quad q_m(0) = 1 .$$

These rationals / if they exist / are unique and are called Padé approximations of order (n, m) . The double infinite array of $p_n(x)/q_m(x)$ $((n=k, m=0,1,2,\dots)\ k=0,1,2,\dots)$ is the Padé table to $f(x)$. If instead of monomials x^k we use any other functions / e.g. orthogonal polynomials / we call the rational functions derived in this way the generalized Padé approximations [2] .If we work with polynomials we cannot hope to get good approximations in problems with an infinite interval. In these cases we get the approximations by annihila - ting as far as possible the leading term in

$$q_m(x) \cdot f(x) - p_n(x) .$$

This is the second method of generalized Padé approximation.

With regard to attainable errors it is generally believed that the rational approximations are essentially no better than the polynomial approximations / see negative theorems [3],[4] /. Experience , nevertheless , seems to show for practical functions / which are quite smooth / that the rational approximations are in fact somewhat superior.

The aim of this paper is to present the strength of the generalized Padé approximation to some special functions in explicit form. Such explicit soluble approximations do not occur in the literature.

In Sections 2 and 3 we give the explicit approximations to $\ln(1+x)$ and $\arctan(x)$, respectively. In Section 4 we consider the function $x^{1/2}$ for $(0,1)$. In Section 5 we get the generalized Padé table to e^{-x} for $(0,\infty)$. In this table the principal diagonal approximants satisfy the relation

$$\max_{0\le x\le\infty}\left|e^{-x}-\frac{P_n(x)}{Q_n(x)}\right| = O\left(\frac{1}{4^n}\right), \quad n\to\infty .$$

There exist in the table other approximants for which

$$\max_{0\le x\le\infty}\left|e^{-x}-\frac{P_n^*(x)}{Q_n^*(x)}\right| = O\left(\frac{1}{q^n}\right), \; q=8.962\ldots, \quad n\to\infty,$$

where $m\to\infty$ too, in such a manner that

$$\lim_{n\to\infty}\frac{m(n)}{n} = \beta = 1.174\ldots ,$$

and the number β is the root of the following equation

$$2\cdot(\beta-1)^{\beta-1}(\beta+2)^{\beta+2} = 3^{2\beta+1}\beta^{2\beta} .$$

It is conjectured [5] that the best approximations in the sense of the minimax norm is

$$\min_{p_n,q_n}\max_{0\le x\le\infty}\left|e^{x}-\frac{P_n^*(x)}{q_n^*(x)}\right| = O\left(\frac{1}{9^n}\right), \quad n\to\infty .$$

Our previous result is near to this conjecture.

2. The logarithm function. We will use its Padé approximation

$$\ln(1+x) = \frac{P_n(x)-S_n(x)}{q_n(x)}, \quad n=1,2,\ldots,$$

where $q_n(x) = {}_2F_1(-n,-n;-2n;-x)$, $S_n(x) = \frac{n!^4 x^{2n+1}}{(2n)!(2n+1)!}\,{}_2F_1(n+1,n+1;2n+2;-x)$,

$$P_n(x) = 2\sum_{j=1}^{n}\frac{(-n)_j^2}{1!(-2n)_j}\sigma_j(-x)^j, \quad \sigma_j = \sum_{l=0}^{j-1}\frac{1}{n-l} .$$

For $n=1$ the approximant is

$$\frac{2x}{2+x} .$$

Now to get the generalized Padé approximations we take the substitutions $x \to \varepsilon e^{it}$, $x \to \varepsilon \bar{e}^{it}$ and sum the terms. The remainder of the Padé approximations has the form for $x \to 0$

$$\frac{S_n(x)}{q_n(x)} = \sum_{k=2n+1}^{\infty} b_k x^k .$$

For $0 \le t \le \pi$ in variable $x/\cos t = 2x-1$, $0 \le x \le 1$ / the remainder term of the generalized Padé approximation is

$$\frac{S_n^*(x)}{Q_n(x)} = \frac{S_n(\varepsilon e^{it})}{q_n(\varepsilon e^{it})} + \frac{S_n(\varepsilon \bar{e}^{it})}{q_n(\varepsilon \bar{e}^{it})} = 2 \sum_{k=2n+1}^{\infty} \varepsilon^k b_k^* T_k^*(x) ,$$

here ε is a suitable parameter $(|\varepsilon|<1)$. Taking into account the simple equality

$$\ln(1+\varepsilon e^{it}) + \ln(1+\varepsilon \bar{e}^{it}) = \ln(1+\varepsilon^2+2\varepsilon\cos t) = \ln(1-\varepsilon)^2 + \ln\left(1+\frac{4\varepsilon}{(1-\varepsilon)^2}x\right)$$

we get the generalized Padé approximation

$$\ln(1+\alpha x) = 2 \ln \frac{1+\sqrt{1+\alpha}}{2} + \frac{P_n(x) - S_n^*(x)}{Q_n(x)} .$$

Here α is another parameter

$$\alpha = \frac{4\varepsilon}{(1-\varepsilon)^2} \qquad \text{or} \qquad \varepsilon = \frac{\alpha}{(1+\sqrt{1+\alpha})^2} .$$

For $\alpha=1$, $\varepsilon = 3-2\sqrt{2}$ the error functions are in Fig 1. Line 1 is the error of the Padé approximation

$$\frac{2x}{2+x}$$

and Line 2 is the absolute value of the error of the generalized Padé approximation

$$2 \ln \frac{1+\sqrt{2}}{2} + 4\varepsilon \frac{\varepsilon - 2 + 4x}{(\varepsilon-2)^2 + 8\varepsilon x} .$$

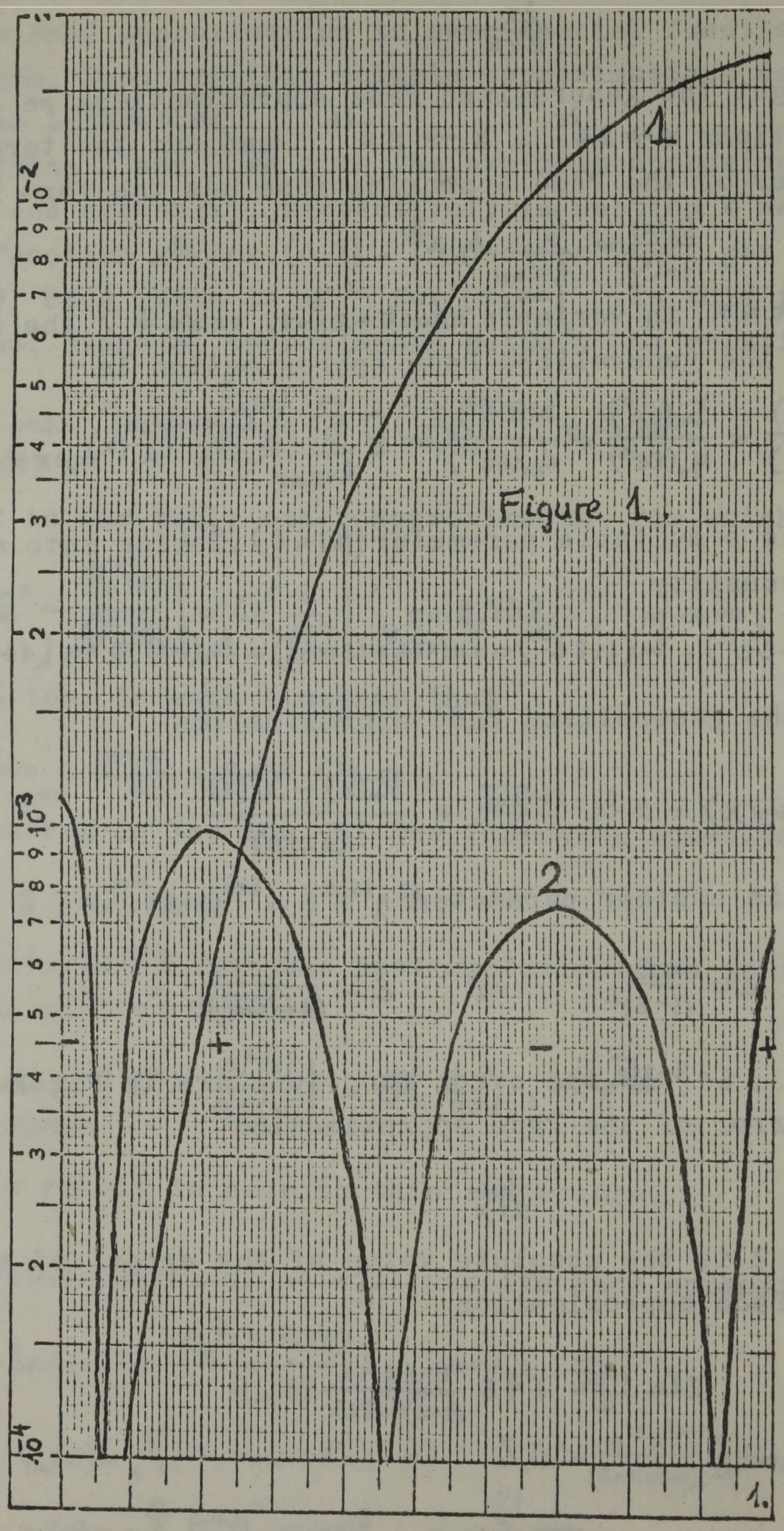


Figure 1.

The absolute value of the maximal error for the generalized Padé approximation is smaller, essentially. We will show this property of the approximation to be valid in general.

To investigate the asymptotic form of the error for $n \to \infty$ we adopt the following equalities from the theory of the hypergeometric functions

$$ {}_2F_1(-n,-n;-2n;-z) = \left(\frac{1+\sqrt{1+z}}{2}\right)^{2n+1}(1+z)^{-\frac{1}{4}}\,{}_2F_1\left(\tfrac{1}{2},\tfrac{1}{2};-n+\tfrac{1}{2};-\frac{z^2}{4\sqrt{1+z}\left(1+\sqrt{1+z}\right)^2}\right), $$

$$ {}_2F_1(n+1,n+1;2n+2;-z) = \left(\frac{2}{1+\sqrt{1+z}}\right)^{2n+1}(1+z)^{-\frac{1}{4}}\,{}_2F_1\left(\tfrac{1}{2},\tfrac{1}{2};n+\tfrac{3}{2};-\frac{z^2}{4\sqrt{1+z}\left(1+\sqrt{1+z}\right)^2}\right). $$

Thus the error of the Padé approximation is

$$ \frac{S_n(x)}{q_n(x)} \approx \frac{n!^4}{(2n)!(2n+1)!}\left[\frac{x}{1+\sqrt{1+x}}\right]^{2n+1}\left\{1+O\left(\tfrac{1}{n}\right)\right\} \quad ; n \to \infty . $$

But for the error of the generalized Padé approximation we get

$$ \frac{S_n^*(x)}{Q_n(x)} \approx \frac{n!^4}{(2n)!(2n+1)!}\left\{\left[\frac{\varepsilon e^{it}}{\left(1+\sqrt{1+\varepsilon e^{it}}\right)^2}\right]^{2n+1}\left(1+O\left(\tfrac{1}{n}\right)\right)+\left[\frac{\varepsilon e^{-it}}{\left(1+\sqrt{1+\varepsilon e^{-it}}\right)^2}\right]^{2n+1}\left(1+O\left(\tfrac{1}{n}\right)\right)\right\} $$

and

$$ \max_{0\le x\le 1}\left|\frac{S_n^*(x)}{Q_n(x)}\right| = 2\,\frac{n!^2}{\left(\frac{1}{2}\right)_n\left(\frac{3}{2}\right)_n}\left[\frac{|\varepsilon|}{\left(1+\sqrt{1-|\varepsilon|}\right)^2}\right]^{2n+1}\left(1+O\left(\tfrac{1}{n}\right)\right) . $$

3.The arctan(x) function . Here we mention briefly a similar result to the previous findings. Applying again a simple idea it is possible to get the generalized Padé approximations in explicit form to the arctan(x) function for $|x| \le 1$. Because of the equality

$$ \arctan(u)+\arctan(v)=\arctan\left(\frac{u+v}{1-u\cdot v}\right) \quad , |u\cdot v|<1, $$

taking $u = \varepsilon e^{it}$, $v = \varepsilon e^{-it}$,

$\cos t = x$, $0 \le t \le \pi$, $-1 \le x \le 1$,

we can get the generalized Padé approximations from the Padé approximations. Namely,

$$\arctan(\alpha x) = \frac{P_n(x) - S_n^*(x)}{Q_n(x)} , \quad \alpha = \frac{2\varepsilon}{1-\varepsilon^2} , \quad \varepsilon = \frac{\alpha}{1+\sqrt{1+\alpha^2}} ,$$

where

$$\frac{P_n(x)}{Q_n(x)} = \frac{p_n(\varepsilon e^{it})}{q_n(\varepsilon e^{it})} + \frac{p_n(\varepsilon \bar{e}^{it})}{q_n(\varepsilon \bar{e}^{it})} ,$$

and p_n and q_n are the Padé numerators and denominators, respectively.

In Fig 2 there are the error functions for $n=1$. Line 1 is the error of the Padé approximation

$$x \cdot \frac{15+4x^2}{15+9x^2}$$

and Line 2 is the absolute value of the error function of the generalized Padé approximation

$$2\varepsilon x \, \frac{1 - \frac{1}{5}\varepsilon^2 + \frac{4}{25}\varepsilon^4 + \frac{16}{15}\varepsilon^2 x^2}{(1-\frac{3}{5}\varepsilon^2)^2 + \frac{12}{5}\varepsilon^2 x^2} , \quad \varepsilon = \sqrt{2} - 1 .$$

4. The function $x^{1/2}$. Next we will show a case of the function having Chebyshev series with slow convergence when the generalized Padé approximation is very far from the best rational approximation.

We consider the function $x^{1/2}$ for $(0 \le x \le 1)$. Its Chebyshev series is

$$x^{\frac{1}{2}} = \frac{1}{\pi}\left(2 - \sum_{k=1}^{\infty} \frac{(-1)^k}{k^2 - 1/4} T_k^*(x)\right) .$$

Our problem is to get the rational approximation in the form

$$x^{\frac{1}{2}} - \frac{P_n(x)}{Q_n(x)} = \sum_{k=2n+1}^{\infty} b_k T_k^*(x) ,$$

where $$P_n(x) = \sum_{j=0}^{n}{}' p_j T_j^*(x) , \quad Q_n(x) = \sum_{j=0}^{n}{}' q_j T_j^*(x) ,$$

here the prime denotes the first term in the sum to be halved. First of all we investigate the following Chebyshev series

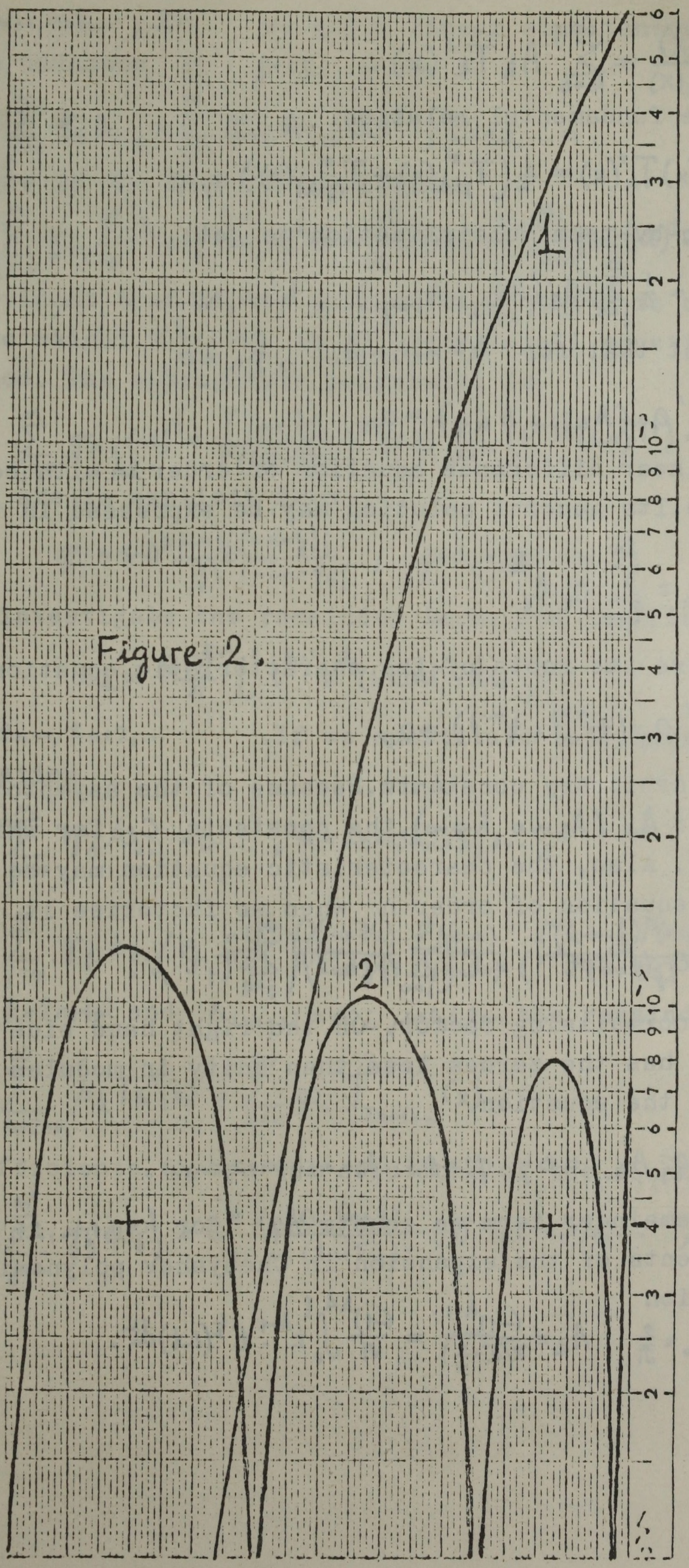


Figure 2.

$$\frac{P_n(x)}{Q_n(x)} = \sum_{k=0}^{\infty}{}' A_k T_k^*(x).$$

Applying the fundamental relation

$$T_j^*(x) T_k^*(x) = \frac{1}{2}\left(T_{j+k}^*(x) + T_{|j-k|}^*(x)\right),$$

using a term-by-term multiplication we get

$$p_k = \frac{1}{2}\sum_{j=0}^{n}{}' q_j (A_{k+j} + A_{|k-j|}), \quad k = 0,1,2,\ldots.$$

It is evident that for $k = n+1, n+2, \ldots$, $p_k = 0$,

$$\sum_{j=0}^{n}{}' q_j (A_{k+j} + A_{k-j}) = 0.$$

This is a linear difference equation of order $2n$ with constant coefficients. Its general solution may be expressed in the form

$$A_k = \sum_{l=1}^{2n} c_l t_l^k$$

where t_l are the roots of the characteristic equation

$$\sum_{j=0}^{n}{}' q_j (t^{n+j} + t^{n-j}) = 0,$$

and c_l are arbitrary constants. Because of the asymptotical expression of A_k ($A_k \to 0$, $k \to \infty$), the constants c_l for which $|t_l| > 1$ must be zero. Also, the characteristic equation may be replaced another equation of order n when we factorise it

$$\sum_{j=0}^{n}{}' q_j (t^{n+j} + t^{n-j}) = C\left(\sum_{j=0}^{n} \gamma_j t^{n-j}\right)\left(\sum_{j=0}^{n} \gamma_j t^j\right).$$

Here C is a normalising factor. On the right hand side the first factor has the zeroes less than unity in modulus. Thus A_k satisfy the equation

$$\sum_{j=0}^{n} \gamma_j A_{k-j} = 0, \quad \gamma_0 = 1, \quad k = n+1, n+2, \ldots.$$

To get the remainder term we equate A_k for $k = 0,1,\ldots,2n$ with the coefficients of the expansion in Chebyshev polynomials to the function

$$A_0 = \frac{2}{\pi}, \quad A_k = \frac{(-1)^{k-1}}{\pi\left(k^2 - \frac{1}{4}\right)} = \frac{(-1)^{k+1}}{\pi} \int_0^1 t^{k-1} \cdot t^{\frac{1}{2}}(1-t)\,dt.$$

For $k=m+1, m+2, \dots, 2n$ we get n linear simultaneous equations for the unknowns γ_j $j=1,2,\dots n$

$$\sum_{j=0}^{m} \gamma_j A_{k-j} = \sum_{j=0}^{m} \gamma_{n-j} A_{j+i} = 0 \quad , \; i=1,2,\dots n .$$

These are in the form of integrals

$$\sum_{j=0}^{m} \gamma_{n-j} \int_0^1 t^{-\frac{1}{2}}(1-t)\, t^{i-1} (-t)^j dt = \int_0^1 t^{-\frac{1}{2}}(1-t)\, t^{i-1} \sum_{j=0}^{m} \gamma_{n-j} (-t)^j dt .$$

It is evident that if the polynomial $\sum_{j=0}^{m} \gamma_{n-j}(-t)^j$ is orthogonal with the weight $t^{-\frac{1}{2}}(1-t)$ to x^{i-1}, $i=1,2,\dots n$ all equations are satisfied. The appropriate polynomial is

$$\sum_{j=0}^{m} \gamma_{n-j} (-t)^j = \frac{\Gamma(m+3/2)}{\Gamma(2n+3/2)}\, t^{1/2}(1-t)^{-1} \frac{d^n}{dt^n}\left(t^{n-\frac{1}{2}}(1-t)^{m+1}\right) .$$

Thus γ_j are in explicit form

$$\gamma_j = (-1)^j \frac{(-m)_j \left(-m+\frac{1}{2}\right)_j}{j!\, (-2n-1/2)_j} \quad , \quad j=0,1,2,\dots n .$$

Now we investigate the remainder term. Its coefficient s_k is

$$s_k = \frac{(-1)^{k-1}}{\pi} \int_0^1 t^{k-1} t^{-\frac{1}{2}}(1-t)\, dt - A_k \, ,$$

for $k = 2n+1, 2n+2, \dots$ .If we substitute this relation into the equation

$$\sum_{j=0}^{m} \gamma_j A_{k-j} = 0 \quad , \; k = 2n+1, 2n+2, \dots ,$$

we get

$$\sum_{j=0}^{m} \gamma_j s_{k-j} = \frac{(-1)^k}{\pi} \frac{\Gamma(m+3/2)}{\Gamma(2n+3/2)} \frac{\Gamma(k-m)}{\Gamma(k-2n)} \cdot \frac{\Gamma(k-n-\frac{1}{2})\Gamma(n+2)}{\Gamma(k+3/2)} .$$

Taking $k=2n+1+i$ and $s_{2n+1+i} = \omega_i$, $i=0,1,2,\dots$ the remainder term is

$$\sum_{k=2n+1}^{\infty} s_k T_k^*(x) = \sum_{i=0}^{\infty} \omega_i T_{2n+1+i}^*(x) .$$

The coefficients ω_i are the solutions of the equation

$$\sum_{j=0}^{m^*} \gamma_j \omega_{i-j} = C^* \frac{(n+1)_i \left(n+\frac{1}{2}\right)_i}{i!\, (2n+5/2)_i} (-1)^i \quad , \; i=0,1,2,\dots \quad ,$$

where

$$n^* = \min(n, i) \quad , C^* = \frac{1}{\pi} \frac{\Gamma(n+1/2)\Gamma(n+1)\Gamma(n+3/2)\Gamma(n+2)}{\Gamma(2n+3/2)\,\Gamma(2n+5/2)} .$$

To get the coefficients ω_i we apply the generating function

$$G(u) = \sum_{i=0}^{\infty} \omega_i u^i .$$

The solution is

$$G(u) = C^* \frac{{}_2F_1(n+1, n+\frac{1}{2}; 2n+\frac{5}{2}; -u)}{{}_2F_1(-n, -n+\frac{1}{2}; -2n-\frac{1}{2}; -u)} .$$

Let us now consider the value of the error at x=o . It is

$$|E_n(0)| = \left| \sum_{i=0}^{\infty} \omega_i T^*_{2n+1+i}(x) \right|_{x=0} = \left| \sum_{i=0}^{\infty} \omega_i (-1)^i \right| = |G(-1)| .$$

But

$$|G(-1)| = \left| C^* \frac{{}_2F_1(n+1, n+\frac{1}{2}; 2n+\frac{5}{2}; 1)}{{}_2F_1(-n, -n+\frac{1}{2}; -2n-\frac{1}{2}; 1)} \right| = \frac{1}{\pi} \frac{1}{(n+\frac{1}{2})(n+1)} .$$

Thus we have proved the following result .The maximal error in the generalized Padé approximation on the interval $(o,1)$ to function $x^{1/2}$ is not less than $O(\frac{1}{n^2})$ $n \to \infty$. This is a sharp negative result .Newman has shown this error in the best approximation to be $O(e^{-c\sqrt{n}})$,where c is a positive constant.

5.Approximation to function e^{-x} in $(0,\infty)$ .Now we proceed by the second method of the generalized Padé approximation.The basic functions will be the Laguerre polynomials.

First we attempt to get the coefficients p_i ; $i=0,1,\dots n$ and q_i $i=0,1,2,\dots n$ in approximation

$$e^{-x} \sim \frac{P_m(x)}{Q_m(x)} = \frac{\sum_{i=0}^{m} p_j L_j(x)}{\sum_{j=0}^{m} q_j L_j(x)} ,$$

by annihilation of the leading terms in

$$S_n^m(x) = e^{-x} Q_m(x) - P_m(x) .$$

The error term of this approximation is

$$\frac{S_n^m(x)}{Q_m(x)} \qquad \text{where} \qquad S_n^m(x) = \sum_{k=n+m+1}^{\infty} b_k L_k(x) .$$

By the orthogonality condition we get

$$p_k = \int_0^\infty e^{-2x} Q_m(x) L_k(x) dx \quad , \; k = 0, 1, 2, \dots n ,$$

$$0 = \int_0^\infty e^{-2x} Q_m(x) L_k(x) dx \quad , \; k = n+1, n+2, \dots, n+m ,$$

$$b_k = \int_0^\infty e^{-2x} Q_m(x) L_k(x) dx \quad , \; k = n+m+1, n+m+2, \dots .$$

First of all we compute an integral which we will use later. Let

$$C_k^m = \int_0^\infty e^{-2x} L_m(x) L_k(x) dx ,$$

and let us define its generating function in two variables

$$G(u_1, u_2) = \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} u_1^k u_2^m C_k^m .$$

A short computation gives

$$G(u_1, u_2) = \frac{1}{(1-u_1)(1-u_2)} \int_0^\infty e^{-2x} \cdot e^{-x\left(\frac{u_1}{1-u_1} + \frac{u_2}{1-u_2}\right)} dx =$$

$$= \frac{1}{2-u_1-u_2} .$$

The integral is

$$C_k^m = \frac{(m+1)_k}{2^{m+k+1} k!} = \frac{1}{2^{m+k+1} k! m!} \int_0^\infty t^{m+k} e^{-t} dt .$$

To compute q_j numbers we employ this integral and the orthogonality property for $k = n+1, n+2, \dots, n+m$,

$$\sum_{j=0}^{m} q_j C_j^k = 0 ,$$

which in another form is

$$\sum_{j=0}^{m} q_j \frac{1}{2^j j!} \int_0^\infty t^{j+k} e^{-t} dt = \int_0^\infty t^k e^{-t} \sum_{j=0}^{m} q_j \frac{1}{j!} \left(\frac{t}{2}\right)^j dt = 0 .$$

When we choose the polynomial

$$\sum_{j=0}^{m} q_j \frac{1}{j!}\left(\frac{t}{2}\right)^j = \frac{(m+1)!}{(n+m+1)!}\, t^{-n-1} e^t \frac{d^m}{dt^m}\left(t^{n+m+1} e^{-t}\right),$$

all our equations are satisfied. From this we get the explicit form of q_j

$$q_j = \frac{(-m)_j}{(n+2)_j}\, 2^j \quad , \; j = 0, 1, 2, \ldots m .$$

A similar computation is needed for p_j and b_k

$$p_j = \sum_{l=0}^{m} q_l c_l^j = \frac{1}{2^{j+1} \cdot j!} \int_0^\infty t^j e^{-t} \sum_{l=0}^{n} q_l \frac{1}{l!}\left(\frac{t}{2}\right)^l dt =$$

$$p_j = \frac{n+1}{2(n+m+1)} \cdot \frac{(-n)_j}{(-n-m)_j} \frac{1}{2^j} \quad , \; j = 0, 1, 2, \ldots n,$$

$$b_{n+m+1+i} = (-1)^m \frac{m!\,(n+1)!}{(n+m+1)!} \frac{1}{2^{n+m+2}} \frac{(m+1)_i}{2^i \cdot i!} \quad , \; i = 0, 1, \ldots .$$

Next, with aid of these explicit results we give some more compact representations for P_n , Q_m , S_n^m respectively.

First we consider

$$P_n(x) = \frac{n+1}{2(n+m+1)} \sum_{j=0}^{n} \frac{(-n)_j}{(-n-m)_j} \frac{1}{2^j} L_j(x)$$

$$= \frac{m!\,(n+1)!}{2(n+m+1)!} \sum_{j=0}^{n} \frac{(m+1)_{n-j}}{(n-j)!} \frac{1}{2^j} L_j(x) .$$

Applying the generating function of Laguerre polynomials we get an integral representation

$$P_n(x) = \frac{m!\,(n+1)!}{(n+m+1)!} \frac{1}{2\pi i} \oint_{|u|=\varepsilon} \frac{1}{u^{n+1}(1-u)^{m+1}} \frac{e^{-x\frac{u}{2-u}}}{2-u}\, du \quad , \; 0 < \varepsilon < 1 .$$

For Q_m a similar formula from the form

$$Q_m(x) = \sum_{j=0}^{m} \frac{(-m)_j}{(n+2)_j} 2^j L_j(x) ,$$

is valid by elementary transformations

$$Q_m(x) = \frac{m!\,(n+1)!}{(n+m+1)!} \frac{1}{2\pi i} \oint_{|u|=\varepsilon} \frac{1}{u^{m+1}(1-u)^{n+1}} \frac{e^{2x\frac{u}{1+u}}}{1+u}\, du .$$

To express S_n^m in an integral form we compute its Laplace transform

$$\int_0^\infty e^{-px} S_n^m(x)dx = \int_0^\infty e^{-px}\left\{\frac{(-1)^m}{2^{n+m+2}}\frac{m!(n+1)!}{(n+m+1)!}\sum_{i=0}^{\infty}\frac{(m+1)_i}{2^i\, i!}L_{n+m+1+i}(x)\right\}dx .$$

Integrating term-by-term then summing the terms we arrive at a very simple expression

$$(-1)^m\frac{m!(n+1)!}{(n+m+1)!}\frac{1}{2^{n+1}}\;\frac{(p-1)^{n+m+1}}{p^{n+1}(p+1)^{m+1}} .$$

Now for S_n^m an integral representation is given by the Mellin formula

$$S_n^m(x) = (-1)^m\frac{m!(n+1)!}{(n+m+1)!}\frac{1}{2^{n+1}}\frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} e^{px}\cdot\frac{(p-1)^{n+m+1}}{p^{n+1}(p+1)^{m+1}}dp .$$

Also, by the Laplace transformation method we can get the following representations

$$P_m(x) = \frac{n+1}{2^{n+1}}\int_0^1 t^m(1+t)^n L_n\left(\frac{1-t}{1+t}x\right)dt ,$$

$$Q_m(x) = (n+1)\int_0^1 (1-t)^n(1-2t)^m L_m\left(-\frac{2t}{1-2t}x\right)dt .$$

From our results we can build the generalized Padé table. But in practice to get the elements of this table it is more efficient to use recursion relations. These relations, difference equations of order three, are as follow

$$4(n+1)(n+m+1)(n+m+2)P_{n+1} = (n+2)(n+m+1)(8n+4m+6-2x)P_n + (n+1)(n+2)(2x-5n-4m-2)P_{n-1} + n(n+1)(n+2)P_{n-2}$$

$$P_0(x) = \frac{1}{2(m+1)} ,\quad P_1(x) = \frac{2m+3-x}{2(m+1)(m+2)} ,\quad P_2(x) = \frac{3}{8}\cdot\frac{4m^2+16m+14-4(m+2)x+x^2}{(m+1)(m+2)(m+3)} ,$$

$$n = 2, 3, 4, \dots ,$$

$$(n+m+1)(n+m+2)Q_{m+1} = (n+m+1)(n-m+2x)Q_m + m(2n+1+m-2x)Q_{m-1} + m(m-1)Q_{m-2}$$

$$Q_0(x) = 1 ,\quad Q_1(x) = \frac{n+2x}{n+2} ,\quad Q_2(x) = \frac{n^2+n+2+4(n-1)x+4x^2}{(n+2)(n+3)} ,$$

$$m = 2, 3, 4, \dots .$$

Next we will investigate the error in the generalized Padé table. For the case $n=m$, we have listed in Table 1 the maximal errors for $n=1,2,3,4,5,6$. One can see the highest errors occur at $x=\infty, n>2$. At this point it is possible to calculate the exact value of the error, which is $E_m(\infty)=\frac{1}{2^{2n+1}}$.

x_i	$E_1(x_i)$	x_i	$E_2(x_i)$	x_i	$E_3(x_i)$
o.	-.25o	o.	.o313	o.	-.oo481
o.6	.oo488	o.4	-.olo3	o.3	.oo162
2.7	-.oo226	1.5	.oo438	1.1	-.oo082
∞	.125	3.5	-.oo193	2.4	.ooo38
		7.o	.oo145	4.3	-.ooo17
		∞	-.o3125	7.2	.oooo99
				12.o	-.oooloo8
				∞	.oo7813

x_i	$E_6(x_i)$
o.	.oooo128
o.16	-.oooo047
o.6	.oooo028
1.3	-.oooo017
2.2	.oooo1o
3.5	-.ooooo5
5.1	.ooooo2
7.1	-.oooooo11
9.5	.oooooo6
12.5	-.oooooo3
16.5	.oooooo2
21.5	-.oooooo3
29.o	.oooooo8
∞	-.oo122o7

x_i	$E_5(x_i)$
o.	-.oooo941
o.16	.oooo3o1
o.7	-.oooo197
1.5	.oooo11o
2.6	-.oooo057
4.2	.oooo027
6.5	-.oooo011
8.5	.ooooo7
12.o	-.ooooo5
17.o	.ooooo4
23.o	-.ooooo8
∞	.oo4883

x_i	$E_4(x_i)$
o.	.ooo679
o.23	-.ooo236
o.9	.ooo129
1.8	-.ooo68
3.3	.ooo32
5.o	-.oooo14
7.9	.ooooo8
11.9	-.ooooo6
17.5	.ooooo9
∞	-.oo1953

TABLE 1.

More generally for $n\to\infty$ the following relation is valid

$$\lim_{n\to\infty} \max_{0\le x<\infty} |E_n(x)|^{\frac{1}{n}} = \frac{1}{4} .$$

$$2\alpha u(1-u) + (1+u)^2[(1+\beta)u - \beta] = 0,$$

for which $u \to \frac{\beta}{2\alpha}$, $\alpha \to \infty$. Our final result is

$$\lim_{n\to\infty} [Q_m^*(n\alpha)]^{\frac{1}{n}} = \frac{1}{u^\beta(1-u)} e^{2\alpha\frac{u}{1+u}}, \quad \beta = \lim_{n\to\infty} \frac{m}{n}.$$

The asterisk denotes that we neglected the constant term $m!(n+1)!/(n+m+1)!$ because of its appearance in the numerator, too.

For S_n^m a similar procedure was carried out. In the integral

$$S_n^m(x) = (-1)^m \frac{m!(n+1)!}{(n+m+1)!} \frac{1}{2^{n+1}} \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} e^{px} \frac{(p-1)^{n+m+1}}{p^{n+1}(p+1)^{m+1}} dp,$$

the main contribution comes from point p, where p is the root of the equation

$$\alpha p(p^2-1) + (1+2\beta)p + 1 = 0,$$

for p, $p \to \frac{1}{\alpha}$, $\alpha \to \infty$ is valid. We then get

$$\lim_{n\to\infty} \left| S_n^{m*}(n\alpha) \right|^{\frac{1}{n}} = \frac{(1-p)^{1+\beta}}{2p(p+1)^\beta} e^{\alpha p}.$$

Finally the maximal error term is

$$\lim_{n\to\infty} \left| \tilde{E}_{n,m}(n\alpha) \right|^{\frac{1}{n}} = \frac{(1-p)^{1+\beta}}{2p(p+1)^\beta} (1-u) u^\beta e^{\alpha(p-\frac{2u}{1+u})}.$$

Let us denote by H the value on the right hand side

$$H = \frac{(1-p)^{1+\beta}}{2p(1+p)^\beta} (1-u) u^\beta e^{\alpha(p-\frac{2u}{1+u})}.$$

Now we investigate the maximal value of H in α. By direct differentiation with respect to α we get

$$\frac{\partial H}{\partial \alpha} = H\left(p - \frac{2u}{1+u}\right).$$

This means that the maximal value there is, where

$$p = \frac{2u}{1+u}.$$

Eliminating α from the equations for u and p we can

solve the equations explicitly . The results are

$$u = \frac{\beta-1}{3(1+\beta)} \ , \quad p = \frac{\beta-1}{2\beta+1} \ .$$

With these, α is in explicit form too

$$\alpha = \frac{(2\beta+1)^3}{3(\beta-1)(\beta+2)} \ .$$

Finally H is / now in β only /

$$\max_{\alpha} H = H^* = \frac{(\beta-1)^{\beta-1}(\beta+2)^{\beta+2}}{\beta^\beta(\beta+1)^{\beta+1}} \frac{1}{3^{2\beta+1}} \ .$$

Let us now consider the value of the error at $x=0$. The modulus of the error at $x=0$ is maximal if x is small. It is not difficult to see that its value is

$$E_{m,m}(0) = 1 - \frac{\frac{1}{2^{m+1}} \int_0^1 t^m (1+t)^m dt}{\int_0^1 (1-t)^m (1-2t)^m dt} \ .$$

By elementary calculations we get

$$E_{m,m}(0) = \frac{I_2}{I_1 + I_2} \ ,$$

$$I_1 = \int_0^{1/2} (1-t)^m (1-2t)^m dt = O\left(\frac{1}{n}\right) , \quad |I_2| = \left| \int_{1/2}^1 (1-t)^m (1-2t)^m dt \right| = \left\{ \frac{\beta^\beta}{2(\beta+1)^{\beta+1}} \right\}^n \left(1 + O\left(\frac{1}{n}\right)\right) .$$

Our result is as follows

$$\lim_{n\to\infty} \left| E_{n,m}(0) \right|^{\frac{1}{n}} = \frac{\beta^\beta}{2(\beta+1)^{\beta+1}} \ .$$

Next we investigate the rate of the convergence with respect to n terms at points $x=0$ and $x=n\alpha$ .Let us define the functions

$$\nu(\beta) = \frac{\beta^{\frac{\beta}{1+\beta}}}{(1+\beta)\, 2^{\frac{1}{(1+\beta)}}} , \qquad t(\beta) = (H^*)^{\frac{1}{1+\beta}} .$$

The function ν varies from $2^{-3/2}$ to 1 in the interval $1<\beta<\infty$. After differentiating we get

$$\nu'(\beta)/\nu(\beta) = \frac{1}{1+\beta} \ln(1+\beta) + \frac{\ln 2}{(1+\beta)^2} > 0 ,$$

the function ν is a strictly monotonic increasing function.

The function $t(\beta)$ varies from $2^{-\frac{1}{2}}=0.707...$ to $9^{-1}=0.111...$ in the interval $1<\beta<\infty$ .Its logarithmic derivative is

$$\frac{t'(\beta)}{t(\beta)} = -\frac{1}{(1+\beta)^2} \ln \frac{3\beta(\beta+2)}{(\beta-1)^2} < 0 .$$

Thus the function $t(\beta)$ is a strictly monotonic decreasing function . This means that the equation $r(\beta)=t(\beta)$ has a root in the interval $1<\beta<\infty$, which is

$$\beta = 1.17400243....$$

The rate of the best convergence is o.364665305. The same quantity conjectured by Saff and Varga [5] is $\frac{1}{3}=0.333...$

REFERENCES

[1] A.R.Reddy:Recent Advances in Chebyshev Rational Approximations on Finite and Infinite Intervals. J.Appr.Theory 22,/1978/ pp.59-84.

[2] E.W.Cheney:Introduction to Approximation Theory. McGraw-Hill,New York.1966.

[3] R.P.Feinerman,D.J.Newman:Polynomial Approximation. Williams and Wilkins,Baltimore.1974.

[4] H.S.Shapiro:Some Negative Theorems of Approximation Theory. Michigan Math.Journal 11,/1964/pp.211-217.

[5] E.B.Saff,R.S.Varga:Some Open Problems. In "Padé and Rational Approximation" / eds.E.B.Saff,R.S.Varga / Academic Press,New York./1977/.

TOPOLOGICAL STRUCTURE OF THE NON-LINEAR MODE COUPLING MODEL EQUATIONS IN A PLASMA I.

Gy.Páris, Á.Ág, G.Németh

Central Research Institute for Physics, Budapest

АННОТАЦИЯ

Исследуется система двух нелинейных дифференциальных уравнений, описывающих нелинейное взаимодействие типа "волна-волна" в присутствии источников /стоков/. Путем соответствующего линейного преобразования система преобразуется к виду, когда положение особых точек в конечной части пространства не зависит от параметров. Вместе с этим, полученная система по виду совпадает с уже изученной Н. Баутиным |7| системой уравнений. Анализируется качественное поведение решений в зависимости от параметров. С помощью численного интегрирования на ЭВМ приводятся примеры существования устойчивых и неустойчивых предельных циклов.

ABSTRACT

The system of equations describing the non-linear two wave interaction is considered with sources. If the system is transformed suitably, the singularities in the finite area of the plane do not depend on the parameters. Thus, it is similar to the system investigated by Bautin using the method of Poincaré and Liapunov. The properties of the solutions as functions of the parameters are reviewed. Some examples are given, yielded by numerical integrations, for the existence and localization of the stable and unstable limit cycles in the finite area, and typical phase diagrams in the entire plane are presented.

Introduction

For the description of the non-linear wave-wave interaction /see, e.g. Tsytovich [1]/, White et al. [2] discussed this relatively simple non-linear system

$$\frac{dI_0}{dt} = \gamma_0 I_0 - \alpha I_0 I_1 + R + \beta_0 I_1^2$$

$$\frac{dI_1}{dt} = -\gamma_1 I_1 + \alpha I_0 I_1 + S + \beta_1 I_1^2 \qquad (*)$$

Here I_0 is the linearly unstable wave intensity /e.g. occupation number/, I_1 is the linearly damped mode, α is the coupling constant, R and S are the coefficients of the spontaneous emission. The terms $\beta_0 I_1^2$ and $\beta_1 I_1^2$ express the selfenhancement stemming from other waves of the same type /see, e.g. [2]/.

For the range of validity of the system $(*)$ see the Appendix.

As Hasegawa has mentioned [3], the system $(*)$ is able to describe many properties of non-linear systems: non-linear damping, saturation, etc. The system $(*)$ is treated by the multiple time scale method in the work of Anderson [4]. It is obvious that the content of this system is not exhausted and the next step is to investigate its topological structure.

It is impossible to give the solution by elementary functions. The range of existence of I_0 and I_1 is arbitrary even by very small R and S values, so much the more is it important to perform a global survey: the

existence of the solution, the possibility of expansion into series, etc.

It should also be noted that in system $(*)$ the sources may have time dependence too. The equations have not only physical but also biological and other applications /see, e.g. Haken [8]/; in these cases the time dependence may be very strong. We intend to investigate this further in a subsequent article. For the sake of simplicity, constant parameters are supposed and the terms I^2, i.e. $\beta_0 = \beta_1 = 0$, are neglected, even though the method which we use permits the cases $\beta_i \neq 0$ to be treated as well. It is important, however, to give the numerical values of the other parameters, cf. the integral curves, Fig. 5, 6.

Systems of type $(*)$ are easy to combine with other spatially dependent operators, expressing, for example, diffusion, etc. This is a forthcoming task too.

Our modell equation is of second order so its integral curves are representable in a plane. The most exhaustive treatment, applying the method of Poincaré, was carried out by Andronov et al. /[5], [6]/; an excellent recapitulation is given by Bautin and Leontovich [7]. We examine the system on this basis [5]-[7].

Transformation of the system $(*)$

For the sake of convenience, by the parallel displacement of the coordinate system

$$I_0 = \xi + \beta \quad ; \quad I_1 = \eta + \delta$$

we get a modified form

$$\begin{aligned} \dot{\xi} &= -\frac{R}{\beta}\xi - \left(\gamma_1 - \frac{S}{\delta}\right)\eta - \alpha\xi\eta \\ \dot{\eta} &= \left(\gamma_0 + \frac{R}{\beta}\right)\xi - \frac{S}{\delta}\eta + \alpha\xi\eta \end{aligned} \qquad (1)$$

Here the constants are

$$\beta = (\gamma_1\gamma_0 - \alpha(S+R) + \Delta)(2\alpha\gamma_0)^{-1}$$

$$\delta = (\gamma_1\gamma_0 + \alpha(R+S) + \Delta)(2\alpha\gamma_1)^{-1} \qquad (1,a)$$

$$\Delta = \sqrt{[\gamma_1\gamma_0 - \alpha(R+S)]^2 + 4\alpha\gamma_1\gamma_0 R}$$

Thereafter applying the following linear transformations

$$x = -\frac{\sqrt{\Delta}}{k}(\xi+\eta)$$

$$y = -\frac{\gamma_0}{k}\xi + \frac{\gamma_1}{k}\eta \qquad (2)$$

$$z = \sqrt{\Delta}\, t$$

we obtain, for the new variable x, y, z , a new system of equations

$$\frac{dx}{dz} = y$$

$$\frac{dy}{dz} = -x - ay + x^2 - Bxy - Cy^2 \qquad (3)$$

Here the new constants are

$$B = \frac{\gamma_1-\gamma_0}{\gamma_1\gamma_0}\sqrt{\Delta}\ ; \qquad C = \frac{\Delta}{\gamma_1\gamma_0}; \qquad k = \frac{\gamma_1+\gamma_0}{\alpha\gamma_1\gamma_0}\Delta^{3/2}$$

$$a = \frac{\gamma_1-\gamma_0}{2\sqrt{\Delta}}\left[1 - C + \frac{\gamma_1+\gamma_0}{\gamma_1-\gamma_0}\frac{\alpha}{\gamma_1\gamma_0}(R+S)\right] = \frac{\gamma_1-\gamma_0}{2\sqrt{\Delta}}[1-C+K] \qquad (3,a)$$

Let us consider in the square bracket the third term, K:

$$K = \frac{\gamma_1+\gamma_0}{\gamma_1-\gamma_0}\frac{\alpha}{\gamma_1\gamma_0}(R+S)$$

Applying the abbreviations of White et al. [2], we put $\gamma_0 = q$, $\alpha = \gamma_1 = 1$ so the constant

$$\frac{K}{R+S} = \frac{1+q}{(1-q)q}$$

depends upon q only. This dependence is given by Fig. 1. Except for the two extrema, to each $K/(R+S)$ -value belongs two q -values. The extrema are

$$\left.\frac{K}{R+S}\right|_{min.} = (\sqrt{2}+1)^2 \; ; \quad \left.\frac{K}{R+S}\right|_{max.} = \frac{1}{(\sqrt{2}+1)^2}$$

Topology of the phase space

The transformation of the system (*) to the form (3) resultsthat the singularities in the finite area of the plane do not depend upon the parametric values. On the other hand from (3) we easily get a differential equation of second order and so we can apply the methods of solution elaborated for this case, e.g. the well-known averaging procedure. But it is more important that the qualitative methods for the investigation of the integral curves are also known /see, e.g. [5], [6], [7]/, these being very useful for a global survey, viz. determination of the specific zones, etc.

System (3) in the bounded x,y -plane has two singular points in the finite area of the plane. The $(0,0)$ -point is a centre, focus or node point, but $(1,0)$ is always a saddle point. The other singular points are to be determined by Poincaré transformation. The transformation

$$x = 1/z \; ; \quad y = u/z$$

provides the singularities in the Poincaré equator with the exception of the end points of the coordinate. The transition $z \to 0$ brings

$$C u^2 + B u - 1 = 0 ; \qquad u_{1,2} = -\frac{B}{2C} \pm \sqrt{\left(\frac{B}{2C}\right)^2 + 1}$$

with roots of opposite sign. These singularities are complicated. The end points of the y -axis are given by the transformation

$$y = 1/z \ , \quad x = v/z$$

if $z \to 0$, $v \to 0$. We get simple node points unless $C = 0$; this particular case needs individual investigation.

Let us suppose that $a = B = 0$. This case is not always physically irrelevant: we have three independent parameters: q, R and S . Now (3) forms a conservative system having a first integral

$$H(x,y) = \left(-\frac{1}{C}x^2 + y^2 + \frac{1+C}{C^2}x - \frac{1+C}{2C^3}\right)e^{2Cx} - h = \text{const.} \qquad (5)$$

Representing this system of graphs on the Poincaré sphere and projecting the lower part onto the (x,y) -plane we get for $C < 1$, Fig. 2; and for $C > 1$, Fig. 3.

The contact curve of the entire system (3) with the conservative system is given by

$$y^2(-a - Bx) = 0 \qquad (6)$$

The double straight line $y^2 = 0$ is a false contact: the integral curves of (3) in the axis $y = 0$ intersect the curves of the conservative system. In general, it is possible to verify that the integral curves of (3) in time /of positive sign/ turn clockwise. Investigation of the isoclines shows that the integral curves, except for the singular points, always intersect the x -axis in a right angle.

Now let the parameters a, B, C be given. If the contact curve does not intersect the domain given by the closed curves of the conservative system, then a limit cycle does not exist but, as will be seen if the contact curve intersects it, the limit cycle does exist.

The separatrix of the conservative case is given by

$$H(x,y) = H(1,0)$$

On this separatrix the energy,

$$h_s = \frac{C-1}{2C^3} e^{2C} \qquad (7)$$

if $C<1$, h_s is negative. The contact curves may be tangents of the domain of the closed durves, if

$$H\left(-\frac{a}{B},0\right) = H(1,0) \qquad (8)$$

The parameters are to be determined from Eq. (8). A root is obviously $a = -B$.
Both roots are provided by the following equation,

$$x_s^2 + \frac{1+C}{C} x_s + \frac{1+C}{2C^2} = \frac{1-C}{2C^2} e^{2C(x_s+1)} \qquad (9)$$

here x_s is the ratio

$$x_s = a/B$$

$x_s = -1$ is a root of this equation. From $(3,a)$ we can express, with the help of K /in $(R+S)$-scaling/,

$$x_s^2 + x_s + \frac{1 + K/2}{K^2 + 4K + 5} = \frac{(2x_s+1)(x_s - K/2)}{K^2 + 4K + 5} e^{2(1+K)\frac{x_s+1}{2x_s+1}} \qquad (10)$$

If the K-s are in $0 \leq g \leq 1$ /cf. Fig. 1/, for small values of K x_s does not differ to any great extent from 0,70. For large values of K $x_s \approx K/2$ The root $x_s = -1$ naturally remains. Further details are to be found in [7].

Conclusions

The topology of (3) shows that system (*) in the entire (I_0, I_1) -plane has integral curves. These curves are divided into groups by the separatrices - depending upon the properties of the singularities.

The linear part of (3) may give a rising or an attenuating solution temporally - depending upon the sign and numerical value of a . If $|a|<2$ the origo is either unstable or stable focus; on the other hand if $|a|>2$ it is a stable or an unstable node point. Figure 4. shows the values of a for the different parameter values. These values of a are generally very small so they do not change very much, neither by rising nor by attenuating amplitudes. Many cycles take place in a narrow zone of phase plane.

For negative a -values $(|a|<2)$ the integral curve moves away from the focus point though the curve may be very near to it, thus the non-linear terms begin to play a role in (3). For the stable case /positive a / the influence of the non-linear terms means that the "attracting region" of the focus becomes narrower, so only in a limited region /depending upon the parameters/, do the integral curves wind round the focus; beyond this zone they move away. Thus for positive a we may expect unstable limit cycles.

For negative a values an opposing effect of the non-linear terms is also possible if the unstable focus has a limited region; and from inside the integral curves approximate a stable limit cycle, and from outside too. In this way, a stationary oscillation is formed.

All these considerations are valuable in the region of singularities in the finite. The behaviour on the entire plane is represented in Figs. 5 and 6; these graphs are applied for our case and are taken from ref. [7].

The influence of the non-linear terms is represented by Liapunov's focal characteristic coefficient, α_3 /see, e.g. [7]/,

$$\alpha_3 = -\frac{\pi}{4} B(1-C)$$

Depending on the sign of α_3 and on the sign of a, we can distinguish four cases:

1. a $\alpha_3 < 0$, $a \geq 0$: stable focus, no limit cycle
   b $\alpha_3 < 0$, $a < 0$: unstable focus, stable limit cycle
2. a $\alpha_3 > 0$, $a > 0$: stable focus, unstable limit cycle
   b $\alpha_3 > 0$, $a \leq 0$: unstable focus, no limit cycle

If $\alpha_3 = 0$, then we must investigate the α_5 coefficient, etc. In Fig. 4. the values of the parameters are presented so we can estimate the fundamental properties of the model system. Some interesting examples are also given.

Numerical examples

The first integrals of the system of equations (3) are determined by numerical methods using an R 20 /ES 1020/ computer for some variations of the parameters. The method is briefly described in Appendix 2.

There are many possible combinations; their physical realities are not discussed here but two combinations are illustrated which are felt to be sufficiently representative.

Figure 7 shows the case $R = 10^{-2}$, $S = 10^{-3}$, $q = 10^{-1}$. Now $\alpha_3 > 0$, $a > 0$, so around a stable focal point we find an unstable limit cycle. As regards a suitable method for reaching it: we integrate with negative time flow so the integral curves winding clockwise now become anticlockwise. In this manner the integral curves drawn by the plotter wind on the unstable limit cycle from anywhere.

Figure 8 shows integral curves for the parameter values $R = -10^{-2}$, $S = 10^{-4}$, $q = 5 \cdot 10^{-2}$, so $\alpha_3 < 0$ and $a < 0$. The thick line means the stable limit cycle, the thin one represents the integral curves stemming from an external and from an internal point. In Fig. 9, we can see the change of the limit cycle if R and S are unchanged but q varies. This figure shows the original $\{I_0, I_1\}$ system too. In Fig. 10 we give the picture of the displacement of the limit cycles in the original Cartesien coordinate system.

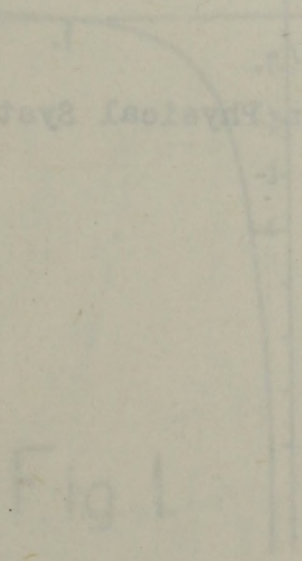


Fig 1

References

[1] V.N. Tsytovich: Nonlinear Effects in Plasmas, Plenum, New York, 1970.

[2] R.B. White, Y.C. Lee, K. Nishikawa: Phys. Rev. Letters, 29, 1315 /1972/

[3] A. Hasegawa: Plasma Instabilities and Nonlinear Effects, Springer, Berlin, 1975.

[4] D. Anderson, A. Bondeson, L. Falk: J. Plasma Phys. 18, /2/, 363 /1977/

[5] A.A. Andronov, E.A. Leontovich, I.I. Gordon, A.G. Maier: Katshestvennaya teoria dinamitsheskih sistem, Nauka, Moscow, 1967.

[6] A.A. Andronov, E.A. Leontovich, I.I. Gordon, A.G. Maier: Teoria bifurkatsii dinamitsheskih sistem na ploskosti, Nauka, Moscow, 1967.

[7] N.N. Bautin, E.A. Leontovich: Metodi i priiomi katshestvennovo issledovania dimanitsheskih sistem na ploskosti, Nauka, Moscow, 1976.

[8] H. Haken: Cooperative Effects, North-Holland, Amsterdam, 1974.

[9] Á. Ág: Comments on the dispersion relation in a magnetoactive turbulent plasma, KFKI report, 1978-21.

[10] A.K. Nekrasov: Nucl. Fusion 16, 113 /1976/.

[11] Chihiro Hayashi: Nonlinear Oscillations in Physical Systems, McGraw-Hill New York, 1964.

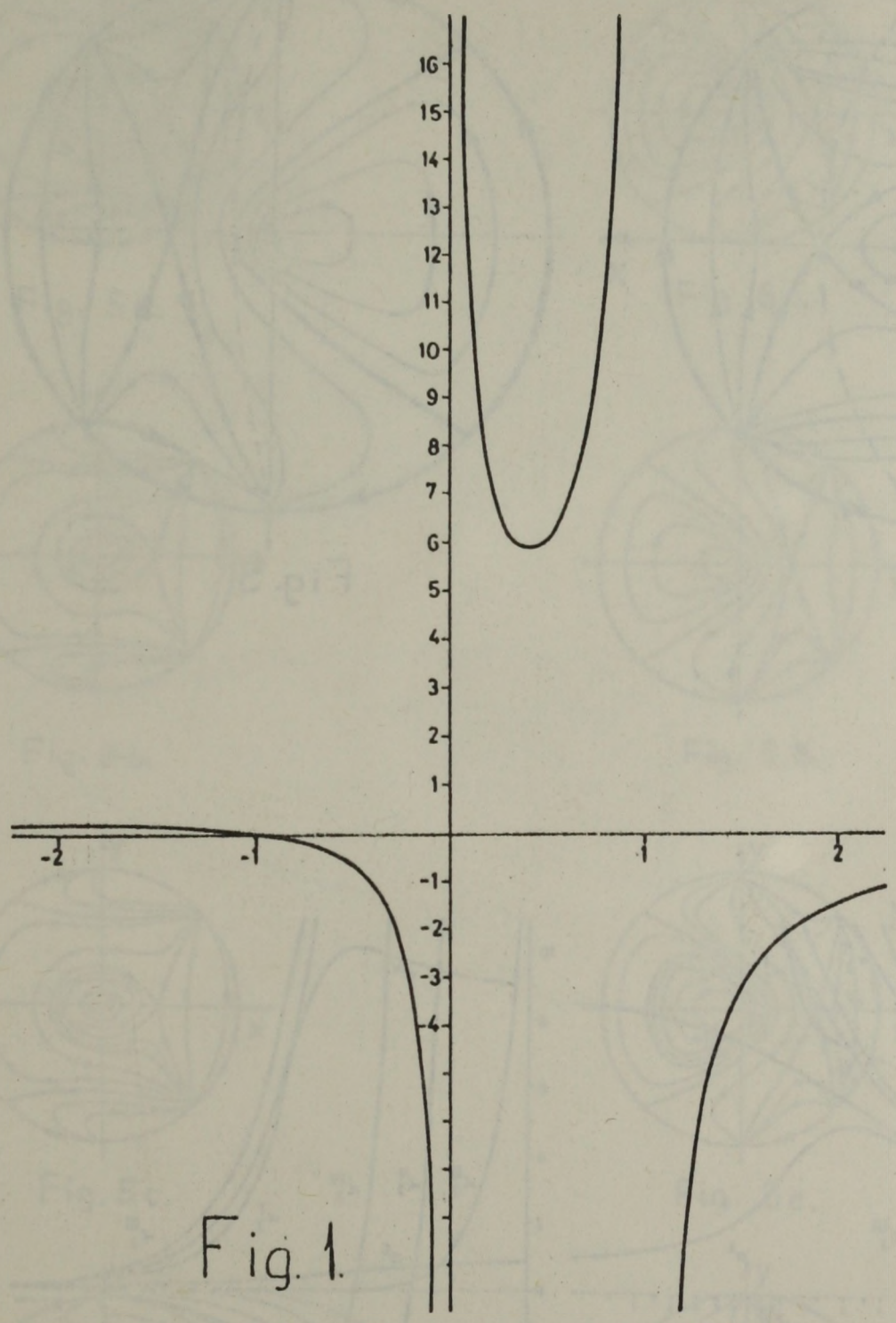


Fig. 1.

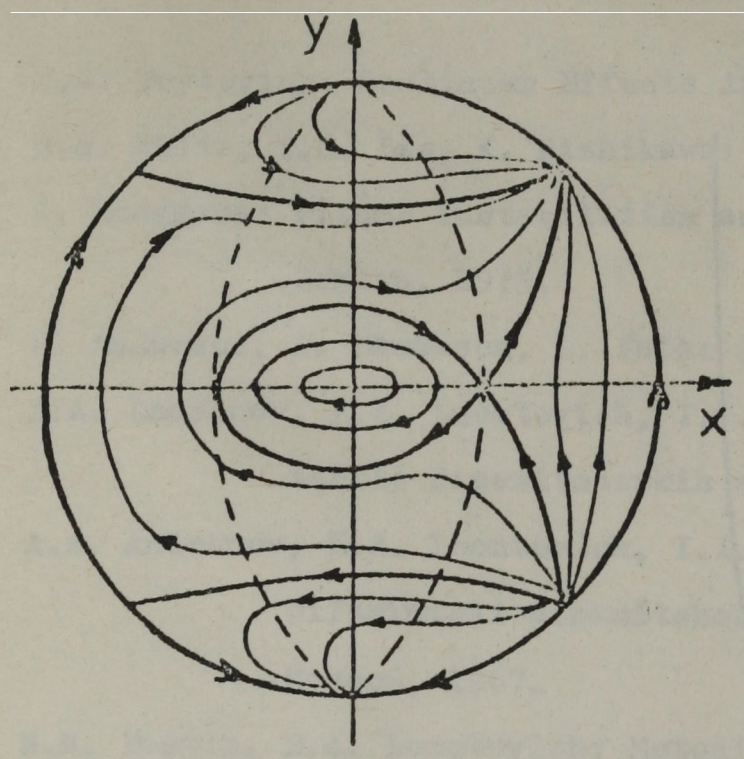


Fig. 2.

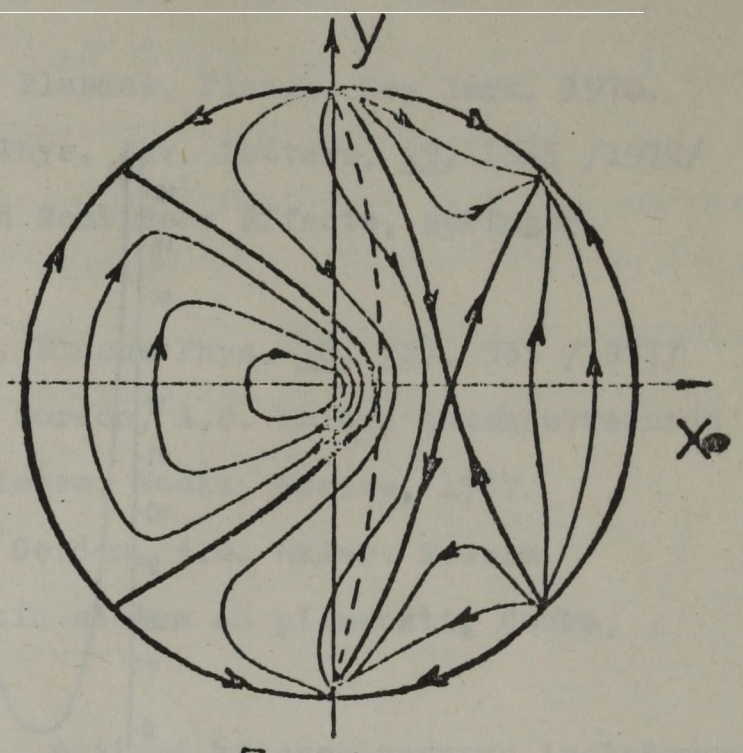


Fig. 3.

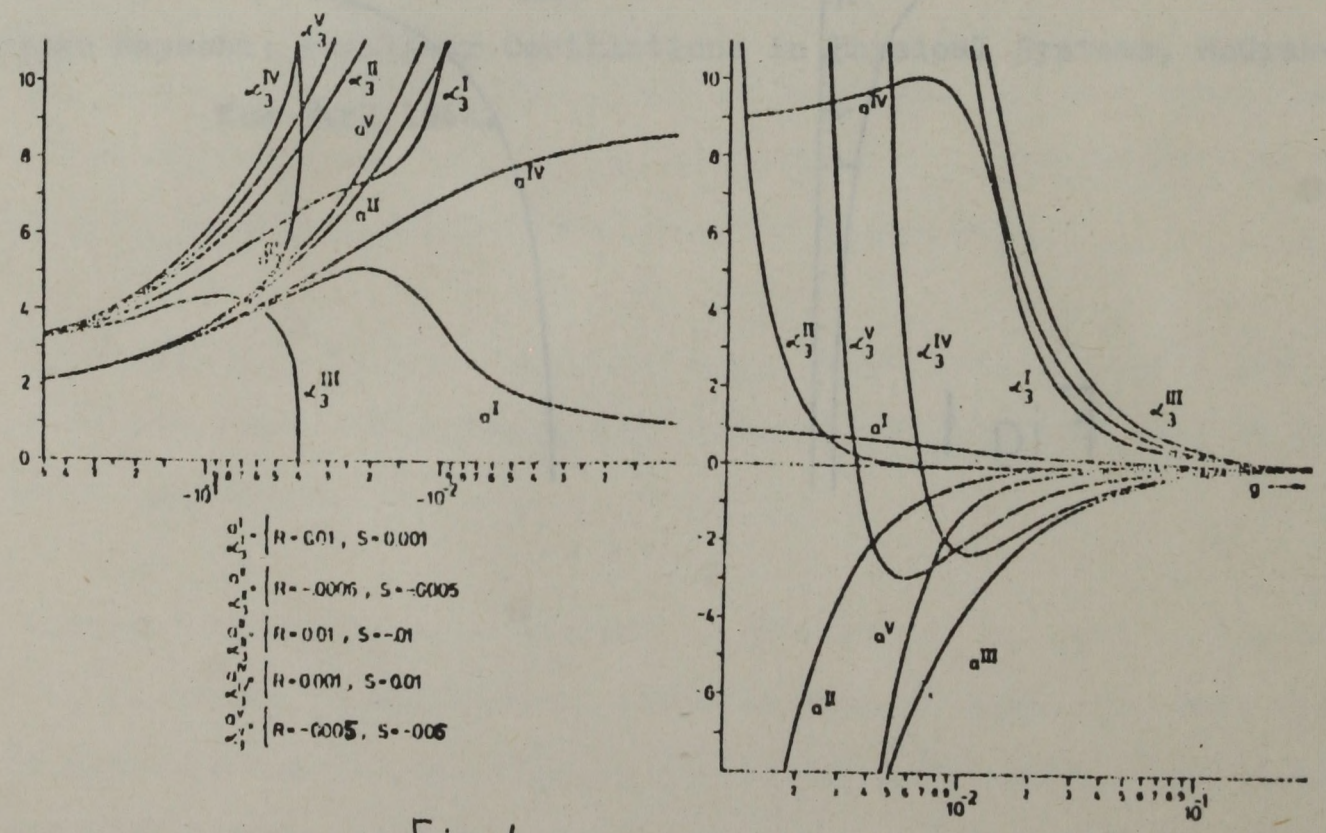


Fig. 4.

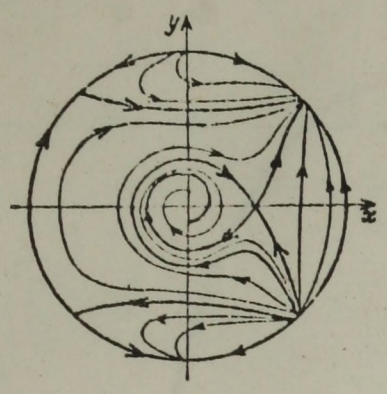


Fig. 5a.

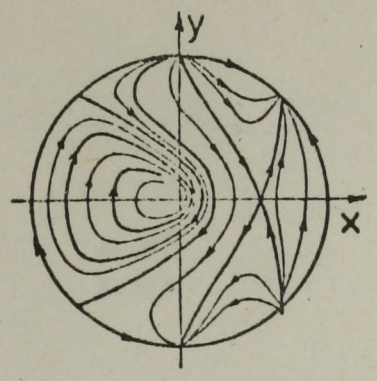


Fig. 6a.

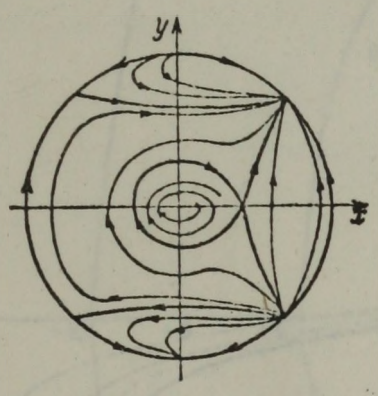


Fig. 5 b.

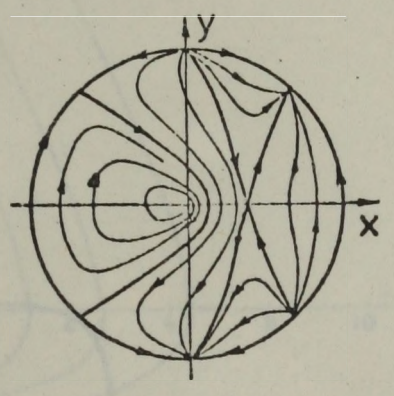


Fig. 6. b.

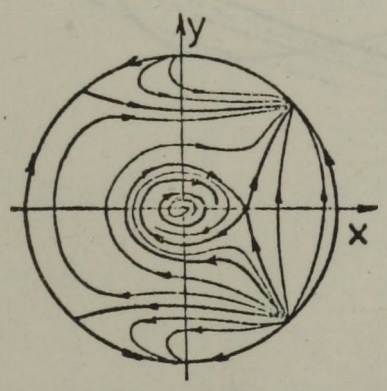


Fig. 5 c.

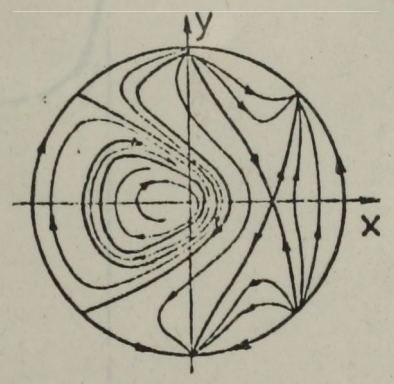


Fig. 6 c.

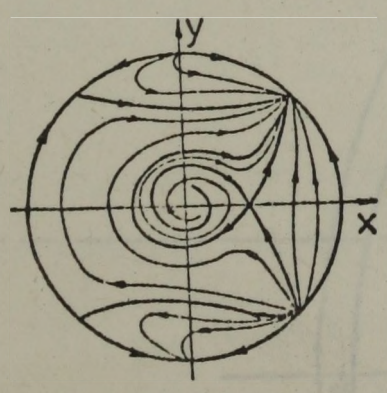


Fig. 5 d.

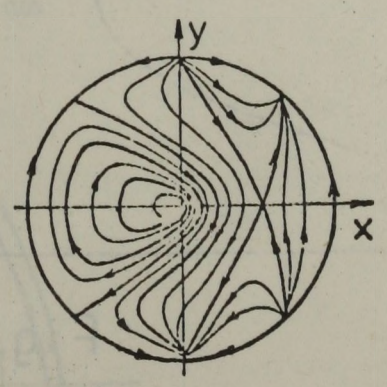


Fig. 6. d.

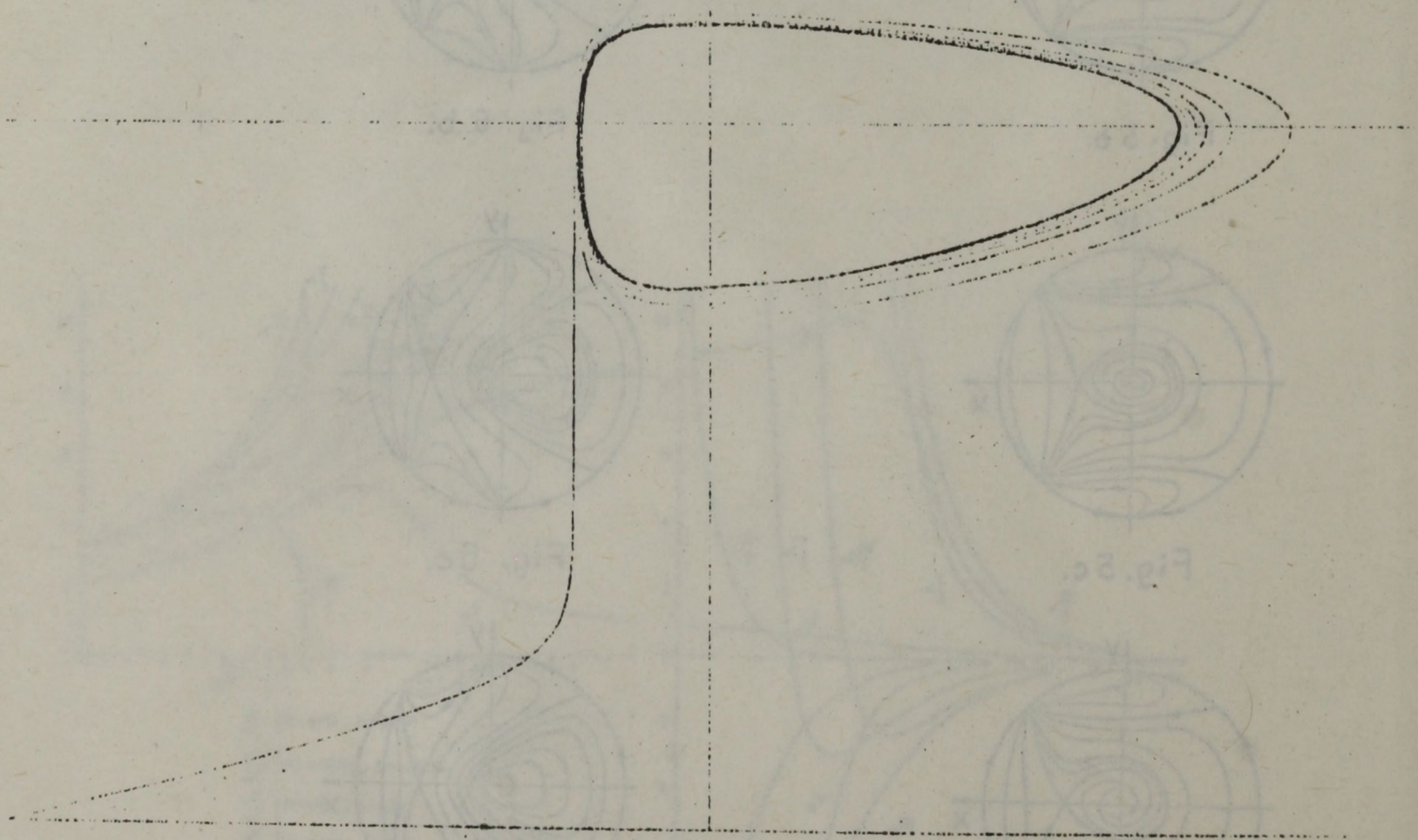


Fig. 7.

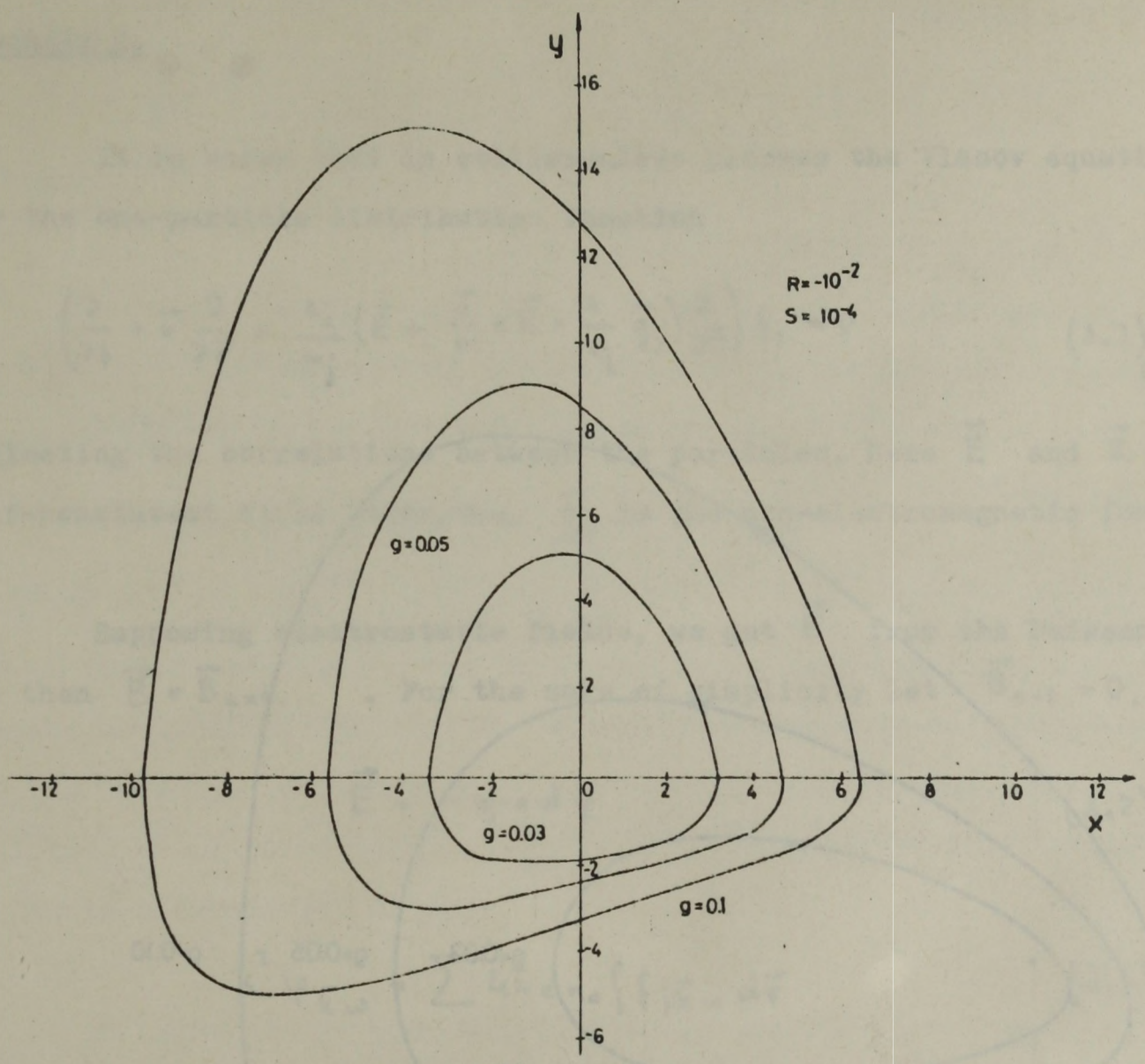

y
16
14
12
10
8
6
4
2
-2
-4
-6
-12
-10
-8
-6
-4
-2
0
2
4
6
8
10
12
x
R= -10^-2
S= 10^-4
g=0.05
g=0.03
g=0.1

Fig. 9.

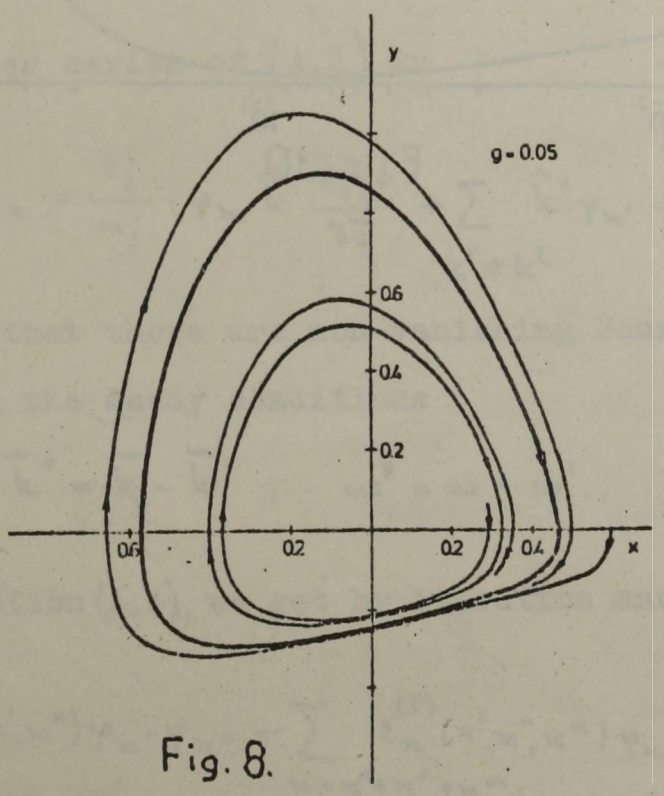

y
g=0.05
0.6
0.4
0.2
0.6
0.2
0.2
0.4
x

Fig. 8.

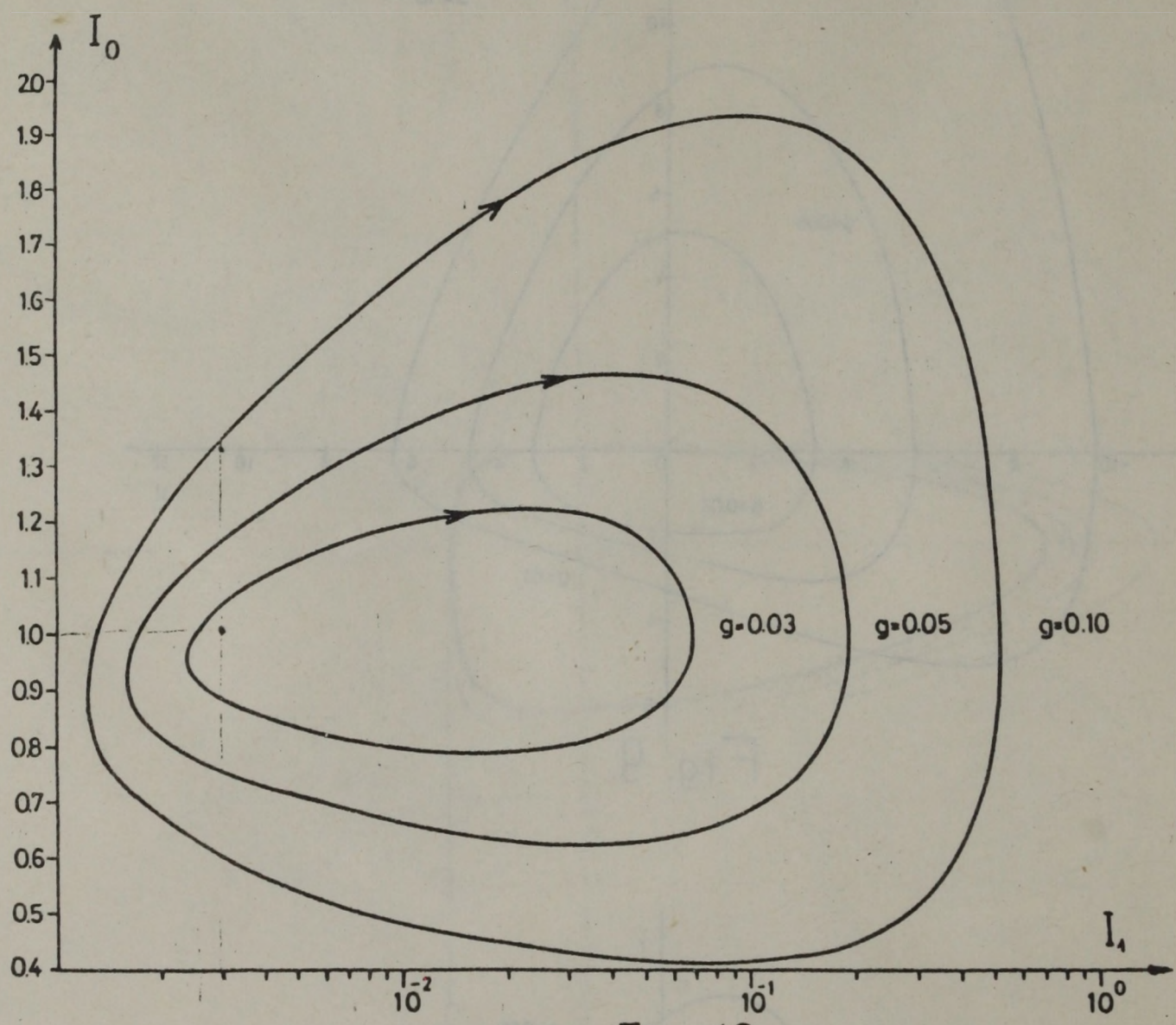

I_0
2.0
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
g=0.03
g=0.05
g=0.10
I_1
10^{-2}
10^{-1}
10^{0}

Fig. 10.

Appendix 1.

It is known that in collisionless plasmas the Vlasov equation holds for the one-particle distribution function

$$\left(\frac{\partial}{\partial t} + \vec{v}\frac{\partial}{\partial \vec{r}} + \frac{e_j}{m_j}\left(\vec{E} + \frac{\vec{v}}{c} \times \vec{B} + \frac{1}{e_j}\vec{q}_j\right)\frac{\partial}{\partial \vec{v}}\right) f_j = 0 \qquad (A.1)$$

neglecting the correlations between the particles. Here $\vec{E}$ and $\vec{B}$ are the self-consistent field strengths, $\vec{q}_j$ is the non-electromagnetic force.

Supposing electrostatic fields, we get $\vec{E}$ from the Poisson equation and then $\vec{B} = \vec{B}_{ext}$. . For the sake of simplicity let $\vec{B}_{ext} = 0$, $\vec{q}_j = 0$

$$\vec{E} = -\operatorname{grad}\varphi \qquad (A.2)$$

$$k^2 \varphi_{\vec{k}\omega} = \sum 4\pi e n_0 \int f_{j\vec{k}\omega}\, d\vec{v} \qquad (A.3)$$

for the Fourier components $\varphi_{\vec{k}\omega}$ and $f_{j\vec{k}\omega}$. We introduce the abbrevation $[\vec{k}, \omega] \doteq \varkappa$

The Fourier series of (A.1) is

$$\left(\vec{k}\vec{v} - \omega\right) f_\varkappa = \frac{e_j}{m_j}\left(\varphi_\varkappa \vec{k}\frac{\partial f_{j0}}{\partial \vec{v}} + \sum_{k'' \neq k'} \vec{k}' \varphi_{\varkappa'} \frac{\partial f_{j\varkappa''}}{\partial \vec{v}}\right. \qquad (A.4)$$

Here we supposed that there are non-vanishing Fourier components for the $\varkappa''$ -terms satisfying the decay conditions

$$\vec{k}'' = \vec{k} - \vec{k}' \; ; \quad \omega'' = \omega - \omega'$$

From equation (A.4), we get by iteration and by the Poisson equation,

$$\varepsilon^{(1)}_\varkappa \varphi_\varkappa + \sum_{\varkappa = \varkappa' + \varkappa''} \varepsilon^{(2)}_\varkappa(\varkappa', \varkappa'')\, \varphi_{\varkappa'} \varphi_{\varkappa''} + \sum_{\varkappa = \varkappa' + \varkappa'' + \varkappa'''} \varepsilon^{(3)}_\varkappa(\varkappa', \varkappa'', \varkappa''')\, \varphi_{\varkappa'} \varphi_{\varkappa''} \varphi_{\varkappa'''} + \ldots = 0 \qquad (A.5)$$

Here the coefficients $\varepsilon^{(i)}$ are the iterated versions of the linear dielectric constants $\varepsilon^{(1)}$. For example, $\varepsilon^{(2)}$ is

$$\varepsilon^{(2)}_{\varkappa}(\varkappa',\varkappa'') = -\frac{1}{2}\sum_j \frac{\omega_{pj}}{k^2}\frac{e_j}{m_j}\int d\vec{v}\,\frac{1}{\omega_{k'}-\vec{k}'\vec{v}+i0}\left(\vec{k}'\frac{\partial}{\partial\vec{v}}\,\frac{1}{\omega_{k''}-\vec{k}''\vec{v}+i0}\,\vec{k}''\frac{\partial}{\partial\vec{v}}+\right.$$

$$\left.+\vec{k}''\frac{\partial}{\partial\vec{v}}\,\frac{1}{\omega_{k'}-\vec{k}'\vec{v}+i0}\,\vec{k}'\frac{\partial}{\partial\vec{v}}\right) f_j^0 \qquad (A.7)$$

This series converges in the weakly coupled plasma theory. In this frame the coupled modes deviate from the linear modes by frequency and wave number shifts so it is possible to write

$$\varepsilon^{(1)}_{\varkappa} = \mathrm{Re}\,\varepsilon^{(1)}(\varkappa_0) + i\,\mathrm{Im}\,\varepsilon^{(1)}(\varkappa_0) + \left(\frac{\partial\varepsilon^{(1)}}{\partial\omega}\right)_{\varkappa_0}\Delta\omega + \left(\frac{\partial\varepsilon^{(1)}}{\partial\vec{k}}\right)_{\varkappa_0}\Delta\vec{k} \qquad (A.8)$$

This is substituted by slowly dependent time and space derivatives, τ, $\vec{\xi}$

$$\varepsilon^{(1)}_{\varkappa}\,\varphi_{\varkappa_0} = i\,\frac{\partial\varepsilon^{(1)}}{\partial\omega_0}\left[-\gamma_0 + \left(\frac{\partial}{\partial\tau} + \vec{v}_g\frac{\partial}{\partial\vec{\xi}}\right)\right]\varphi_{\varkappa_0}(\tau,\vec{\xi}) \qquad (A.9)$$

here $\vec{v}_g = \dfrac{\partial\omega}{\partial\vec{k}}$

$$\gamma_0 = -\frac{\mathrm{Im}\,\varepsilon^{(1)}}{\partial\varepsilon/\partial\omega_0}\;; \qquad \frac{\partial\varepsilon}{\partial\omega_0} = \left(\frac{\partial\,\mathrm{Re}\,\varepsilon^{1}(\varkappa)}{\partial\omega}\right)_{\varkappa=\varkappa_0} \qquad (A.10)$$

Supposing three interacting waves, we can form similar equations for the three potentials $\varphi_{\varkappa_0}$, $\varphi_{\varkappa_1}$ and $\varphi_{\varkappa_2}$

$$i\frac{\partial\varepsilon}{\partial\omega_0}\left(\frac{\partial}{\partial\tau} + \vec{v}_{g0}\frac{\partial}{\partial\vec{\xi}} - \gamma_0\right)\varphi_{\varkappa_0} = -\varepsilon^{(2)}_{\varkappa_0}(\varkappa_1,\varkappa_2)\,\varphi_{\varkappa_1}\varphi_{\varkappa_2} + \theta(\varepsilon^{(3)})$$

$$i\frac{\partial\varepsilon}{\partial\omega_1}\left(\frac{\partial}{\partial\tau} + \vec{v}_{g1}\frac{\partial}{\partial\vec{\xi}} - \gamma_1\right)\varphi_{\varkappa_1} = -\varepsilon^{(2)}_{\varkappa_1}(-\varkappa_2,\varkappa_0)\,\varphi_{-\varkappa_2}\varphi_{\varkappa_0} + \theta(\varepsilon^{(3)}) \qquad (A.11)$$

$$i\frac{\partial\varepsilon}{\partial\omega_2}\left(\frac{\partial}{\partial\tau} + \vec{v}_{g2}\frac{\partial}{\partial\vec{\xi}} - \gamma_2\right)\varphi_{\varkappa_2} = -\varepsilon^{(2)}_{\varkappa_2}(\varkappa_0,-\varkappa_1)\,\varphi_{\varkappa_0}\varphi_{-\varkappa_1} + \theta(\varepsilon^{(3)})$$

Using normed quantities we get a compact system

$$i S_{n_0} \frac{\partial A_{n_0}}{\partial \tau} = \gamma_0' A_{n_0} + V A_{n_1} A_{n_2} \qquad \text{(A.12)}$$

$$\vdots$$

with

$$\vec{k}_0 = \vec{k}_1 + \vec{k}_2, \quad \omega_0 = \omega_1 + \omega_2 \qquad \text{(A.13)}$$

In the incoherent case we sum up those waves which satisfy the decay conditions. If we suppose the third wave strongly damped, we finally get for the spatially uniform case in RPA-approximation

$$\begin{aligned} \dot{I}_0 &= \gamma_0 I_0 - \alpha I_0 I_1 + R \\ \dot{I}_1 &= -\gamma_1 I_1 + \alpha I_0 I_1 + S \end{aligned} \qquad \text{(A.14)}$$

Here we introduced the R and S source terms stemming from the spontaneous emission or external pumping. The coefficients may exhibit slow time dependence too. The numerical calculation of the coefficients in a relatively simple case is already complicated, see e.g. [1], [3], [8], [9]. We hope to treat this problem in a subsequent article.

In higher order approximation there are also terms with I^2 /see e.g. [2], especially for biological systems/.

Appendix 2.

Equation (3), in the form of a second order equation yields

$$\frac{d^2x}{dz^2} + x = -a\frac{dx}{dz} + p\left[x^2 - Bx\frac{dx}{dz} - C\left(\frac{dx}{dz}\right)^2\right] \qquad (A.15)$$

Here p is an arbitrary constant. Let be $x = px'$ and thereafter omit the $'$, we apply the delta method for the numerical integration of (A.15) /see, e.g. [11]/.

This method is based upon the constancy of the right hand side of Equ. (A.15) in a step of the integration. With this value this part of the integral curve is represented by a part of displaced circle in the x-axis.

It is necessary to apply many steps because the curves are generally spirals and many winds need to be used for a good approximation of a limit cycle. In such a manner the standard technique produces considerable errors. By a suitable choice of p, in an empirical manner, it was possible to ensure that the change of the right hand side remained small with each step of the integration.

ЧИСЛЕННЫЙ АНАЛИЗ ПРОСТРАНСТВЕННОГО РАЗВИТИЯ ЛАВИНЫ СВЕРХИЗЛУЧЕНИЯ

А.Н.Тихонов, А.В.Андреев, В.Я.Галкин, Ю.А.Ильинский,
О.Ю.Тихомиров
Московский государственный университет им. М.В.
Ломоносова, Москва

АННОТАЦИЯ

В статье поставлена и численно решается задача описания как временного, так и пространственного изменения плотности поля. Выводятся квантовомеханические уравнения процесса коллективного спонтанного излучения. Для полученной полулинейной гиперболической системы первого порядка ставится смешанная начальная и граничная задача, анализируется ее корректность. При решении используются два метода характеристик численного интегрирования: на основе модифицированного метода Эйлера и - метода трапеций, что приводит соответственно к явной и неявной разностным схемам. Проводится анализ и интерпретация результатов численных расчетов.

ABSTRACT

The problem of time-space variation of e.m. field density during the process of superradiance /SR/ is discussed. Quantum mechanical equations of SR are obtained. One gives a numerical solution of above equations. A mixed inital and boundary problem is posed for quasi-linear hyperbolic system, and its right-pouseness is analysed. Two methods of numerical integraiton are used: in the base of modyfied Euler method and of trapeze method /explicit and implicit differential schemes/. The analysis and interpretation of numerical results is made.

В в е д е н и е. В последнее время все больший интерес привлекает теория коллективного спонтанного излучения для сверхизлучения (СИ). Этот интерес обусловлен главным образом двумя причинами:во-первых, тем, что существующие теории,согласуясь качественно, не дают пока количественного описания эффекта, и, во-вторых, обсуждается возможность наблюдения эффекта в коротковолновом диапазоне:вакуумном УФ, рентгеновском и гамма, где он может явиться единственным механизмом генерации когерентного излучения.

Среди подходов к описанию СИ можно выделить два основных -- квантовый и квазиклассический. Преимущество квантового подхода заключается в том, что он дает описание процессов изотропного спонтанного распада, которые играют существенную роль в развитии сверхизлучающей лавины. Преимущество же квазиклассического подхода, как отмечалось в [I] , состоит в том, что он позволяет учитывать пространственные изменения поля.Однако квазиклассические уравнения не учитывают процессов изотропного спонтанного распада и поэтому требуют соответствующего дополнения. Ясно, что такое введение дополнительных членов никогда нельзя считать однозначным.

В настоящей работе произведен численный анализ квантовых уравнений, учитывающих пространственное изменение плотности поля, на основе разработанных алгоритмов вычислений и программного обеспечения.Для описания электромагнитного поля использованы не операторы рождения и уничтожения квантов в выделенной моде, как это обычно делалось в квантовой теории СИ, а оператор векторного потенциала $\vec{A}$ и обобщенный импульс $\vec{B}$. Это позволило получить уравнения, в которые вместо числа квантов в данной моде входит плотность полевого гамильтониана. Таким образом,поставлена и численно решается задача описания как временного, так и пространственного изменения плотности

поля.

2. К в а н т о в о м е х а н и ч е с к а я с и с т е м а у р а в н е н и й. Гамильтониан системы атомов, взаимодействующих друг с другом посредством когерентного электромагнитного поля, можно записать в виде

$$H = H_f + H_a + H_{int}, \qquad (I)$$

где H_f -гамильтониан электромагнитного поля, H_a -гамильтониан атомной подсистемы, H_{int} -гамильтониан взаимодействия.

Гамильтониан H_f , выраженный через оператор векторного потенциала $\vec{A}(\vec{r},t)$, зависящего от радиуса - вектора $\vec{r}$ и времени t , и канонически сопряженный ему обобщенный импульс $\vec{B}(\vec{r},t)$, равный

$$\vec{B}(\vec{r},t) = \frac{1}{4\pi c^2}\frac{\partial A(r,t)}{\partial t}, \qquad (2)$$

имеет следующий вид

$$H_f = \int\left[2\pi c^2\vec{B}^2 + \frac{1}{2\pi}(rot\,\vec{A})^2\right]dV. \qquad (3)$$

Операторы $\vec{A}$ и $\vec{B}$ удовлетворяют следующим коммутационным соотношениям [2]

$$[A_\alpha(\vec{r},t), A_\beta(\vec{r}',t)] = [B_\alpha(\vec{r},t), B_\beta(\vec{r}',t)] = 0, \qquad (4)$$

$$[A_\alpha(\vec{r},t), B_\beta(\vec{r}',t)] = ih\delta_{\alpha\beta}\delta(\vec{r}-\vec{r}'),$$

где α , β = I,2,3 ; $\delta_{\alpha\beta}$ - символ Кромекера, $\delta(.)$- δ-функция Дирака.

Не конкретизируя среду, запишем гамильтониан взаимодействия в следующем виде

$$H_{int} = -\frac{1}{c}\int\vec{j}(\vec{r})\vec{A}(\vec{r})dV, \qquad (5)$$

где $\vec{j}(\vec{r})$ -плотность тока перехода.

Используя уравнения движения для операторов в гайзенберговском представлении с помощью коммутационных соотношений (4) и гамильтонианов (I),(3),(5), несложно получить следующие уравнения

$$\frac{\partial\vec{A}(\vec{r},t)}{\partial t} = 4\pi c^2\vec{B}(\vec{r},t), \qquad (6)$$

$$\frac{\partial\vec{B}(\vec{r},t)}{\partial t} = \frac{1}{4\pi}rot\,rot\,\vec{A}(\vec{r},t) + \frac{1}{c}\vec{j}(\vec{r},t),$$

а выражение для скорости изменения плотности полевого гамильтониана $\mathcal{H}_f$ ($H_f = \int \mathcal{H}_f dV$) имеет вид

$$\frac{\partial \mathcal{H}_f}{\partial t} = -c^2/2\left(\vec{B}\,\mathrm{rot}\,\mathrm{rot}\,\vec{A} + \mathrm{rot}\,\mathrm{rot}\,\vec{A}\vec{B}\right) + \tag{7}$$
$$+ c^2/2\left(\mathrm{rot}\,\vec{B}\,\mathrm{rot}\,\vec{A} + \mathrm{rot}\,\vec{A}\,\mathrm{rot}\,\vec{B}\right) + 4\pi c\vec{j}\vec{B}.$$

Решение системы (6) будем искать в виде

$$\vec{A}(\vec{r},t) = \vec{A}^+(\vec{r},t)e^{i(\omega t - \vec{k}_0\vec{r})} - \vec{A}^-(\vec{r},t)e^{-i(\omega t + \vec{k}_0\vec{r})}, \tag{8}$$
$$\vec{B}(\vec{r},t) = \frac{i\omega}{4\pi c^2}\left[\vec{A}^+(\vec{r},t)e^{i(\omega t - \vec{k}_0\vec{r})} - \vec{A}^-(\vec{r},t)e^{-i(\omega t + \vec{k}_0\vec{r})}\right],$$

где волновой вектор $\vec{k}_0$, по модулю равный $k_0 = |\vec{k}_0| = \frac{\omega}{c}$, направлен вдоль выделенной оси образца (далее будем обозначать ее z), а амплитуды $\vec{A}^+$ и $\vec{A}^-$ удовлетворяют условиям

$$\left|\partial/\partial t\langle\vec{A}^\pm\rangle\right| \ll \omega\langle\vec{A}^\pm\rangle, \quad \left|\partial/\partial t\langle\vec{A}^\pm\rangle\right| \ll k_0\langle\vec{A}^\pm\rangle, \tag{9}$$

где $\langle . \rangle$ -квантовомеханическое среднее.

Рассматривая среду как ансамбль из N двухуровневых атомов после подстановки разложения (8) в гамильтонианы (3), (5) и усреднения их по периоду $T = \frac{2\pi}{\omega}$, получим

$$H = \frac{\hbar\omega_0}{2}\sum_{\alpha=1}^{N}\sigma_z^{(\alpha)} - \frac{1}{c}\sum_{\alpha=1}^{N}\Big|\vec{m}\sigma_+^{(\alpha)}e^{i\vec{k}_0\vec{r}_\alpha}\vec{A}^-(\vec{r}_\alpha,t) + \tag{10}$$
$$+ \vec{m}^*\sigma_-^{(\alpha)}e^{-i\vec{k}_0\vec{r}_\alpha}\vec{A}^+(\vec{r}_\alpha,t)\Big],$$

где $\hbar\omega_0$ - энергия перехода атома, а матричный элемент плотности тока представлен в виде

$$\langle +\left|\hat{\vec{j}}(r_\alpha,t)\right|-\rangle = \vec{m}\,\delta(\vec{r}-\vec{r}_\alpha),$$

+ или - соответствует возбужденному и основному состояниям.

Используя правила коммутации операторов Паули $\sigma_+, \sigma_-, \sigma_z$:

$$\left[\sigma_+^{(\alpha)}, \sigma_-^{(\beta)}\right] = \sigma_z^{(\alpha)}\delta_{\alpha\beta}, \quad \left[\sigma_\pm^{(\alpha)}, \sigma_z^{(\beta)}\right] = \mp 2\sigma_\pm^{(\alpha)}\delta_{\alpha\beta},$$

получим следующую операторную систему уравнений, описывающих динамику системы "атомы + поле",

$$\frac{\partial\hat{n}}{\partial t} + c\frac{\partial\hat{n}}{\partial z} = \frac{i}{\hbar c}\left(m^*A^+R^- - mR^+A^-\right), \tag{11}$$

$$\frac{\partial}{\partial t}A^+R^- + c\frac{\partial}{\partial z}A^+R^- = i(\omega-\omega_0)A^+R^- -$$
$$-i\frac{2\pi c}{\omega}m\left(\hat{n}\hat{R}_z + R^+R^-\right),$$

$$\frac{\partial}{\partial t}A^-R^+ + c\frac{\partial}{\partial z}A^-R^+ = -i(\omega-\omega_0)A^-R^+ + i\frac{2\pi c}{\omega}m^*(\hat{n}\hat{R}_z + R^+R^-),$$

$$\frac{\partial}{\partial t}R^+R^- = \frac{i}{\hbar c}\left(m^*A^+\hat{R}_zR^- - mR^+\hat{R}_zA^-\right),$$

$$\frac{\partial}{\partial t}\hat{R}_z = -2\frac{i}{\hbar c}\left(m^*A^+R^- - mR^+R^-\right),$$

где $\hat{n} = \frac{1}{\hbar\omega}\mathcal{H}_f$ – оператор плотности квантов,

$$R^{\pm} = \frac{1}{V_1}\sum_{\alpha\in V_1}\sigma_{\pm}^{(\alpha)}e^{\pm(\vec{k}_0\vec{r}_\alpha-\omega t)},$$

$$\hat{R}_z = \frac{1}{V_1}\sum_{\alpha\in V_1}\sigma_z^{(\alpha)}, \qquad m = \sum_{\nu=1}^{2}\vec{e}^{(\nu)}\vec{m},$$

$\vec{e}^{(\nu)}$ –единичные вектора поляризации. При выводе (II) произведено усреднение операторных уравнений по объему $V_1 = V_1(\vec{r})$ размером, много большим длины волны излучения λ, но меньшим характерной длины изменения амплитуд $\vec{A}^+$, $\vec{A}^-$.

Наконец, усредняя систему операторных уравнений (II) с помощью начальной матрицы плотности и вводя релаксационные члены, получим

$$\frac{\partial n}{\partial t} + c\frac{\partial n}{\partial z} + \frac{n}{\tau} = F,$$

$$\frac{\partial F}{\partial t} + \frac{1}{2}\left(\frac{1}{\tau}+\frac{1}{T_2}\right)F + c\frac{\partial F}{\partial z} = \frac{1}{T_0^2}(nR_z + S_0 + S_1), \qquad (12)$$

$$\frac{\partial S_1}{\partial t} + \frac{S_1}{T_2} = FR_z,$$

$$\frac{\partial R_z}{\partial t} = -2F,$$

где $n = \langle\hat{n}\rangle$, $R_z = \langle\hat{R}_z\rangle$, $\frac{1}{T_0^2} = \frac{4\pi|m|^2}{\hbar\omega}$,

$$F = \frac{i}{\hbar\omega}\left(m^*\langle A^+R^-\rangle - m\langle R^+A^-\rangle\right),$$

$$S_0 = \frac{1}{V_1^2}\sum_{\alpha\in V_1}\langle\sigma_+^{(\alpha)}\sigma_-^{(\alpha)}\rangle, \quad S_1 = \frac{1}{V_1^2}\sum_{\substack{\alpha,\beta\in V_1\\ \alpha\neq\beta}}\langle\sigma_+^{(\alpha)}\sigma_-^{(\beta)}\rangle e^{i\vec{k}_0(\vec{r}_\alpha-\vec{r}_\beta)},$$

$\tau = L/c$, L –длина образца по оси z, T_2– время расфазировки (время поперечной релаксации).

3. Анализ постановки задачи. Переходя к безразмерному времени $t = t/\tau$, вводя пространственную

переменную $z = x/L$ и нормируя неизвестные функции к числу излучателей N, получим следующую систему дифференциальных уравнений в частных производных (при переходе от (I2) к (I3), учтем также волну, бегущую налево)

$$\frac{\partial n_k}{\partial t} - (-1)^k \frac{\partial n_k}{\partial z} = F_k - n_k,$$

$$\frac{\partial F_k}{\partial t} - (-1)^k \frac{\partial F_k}{\partial z} = \frac{\beta}{2}\left[n_k R_z + \frac{S}{2} + \bar{G}(1+R_z)\right] - \frac{1}{2}(1+\alpha_1)F_k, \quad (I3)$$

$$\frac{\partial S}{\partial t} + \alpha_1 S = (F_1 + F_2)R_z,$$

$$\frac{\partial R_z}{\partial t} = -2(F_1 + F_2)$$

($K = 1$ – волна направо, $K = 2$ – волна налево), с начальными условиями

$$n_k(z,0) = F_k(z,0) = S(z,0) = 0, \qquad R_z(z,0) = 1, \quad k = 1,2, \quad (I4)$$

а также с однородными граничными условиями

$$n_1(0,t) = F_1(0,t) = 0, \quad n_2(1,t) = F_2(1,t) = 0. \quad (I5)$$

Здесь $\alpha_1 = \tau/T_2$, $\beta = \tau/T_2 \mu_0 L$, где μ_0 – коэффициент усиления среды, $\bar{G} = \varkappa/N$, где $\varkappa$ – плотность числа мод. Граничные условия определяют отсутствие внешенго сигнала. В начальный момент времени поле находится в вакуумном состоянии, среда полностью инвертирована, состояния излучателей некоррелированы.

Решение ищется в области плоскости (z, t)

$$\Pi = \{[0,1] \times [0,T]\}. \quad (I6)$$

Таким образом ставится смешанная начальная и граничная задача для гиперболической системы (I3) полулинейных дифференциальных уравнений в частных производных I-го порядка, определяющая пространственно-временное развитие излучения. Заметим, что опорными характеристиками являются прямые

$$z = 0, \; z = t, \; z = -t; \quad (I7)$$

причем все характеристики кратные.

На рис. I представлена структура области Π. Подобласть G, образованная пересечением оси абсцисс и характеристик $z = t$ и $z = -t + 1$, представляет собой область определенности задачи Коши для (I3); подобласть G_1, образованная пересечением трех различных характеристик $z = 0$, $z = t$, $z = -t + 1$, такова, что здесь на решение влияние оказывают только граничные условия (заданные на Γ_1 : $z = 0$) в результате распростране-

ния волны, бегущей направо ; для области G_2 , образованной пересечением характеристик $z = 1$, $z = t$ и $z = -t + 1$,справедливо то же утверждение относительно граничной кривой Γ_2 и волны, бегущей налево.

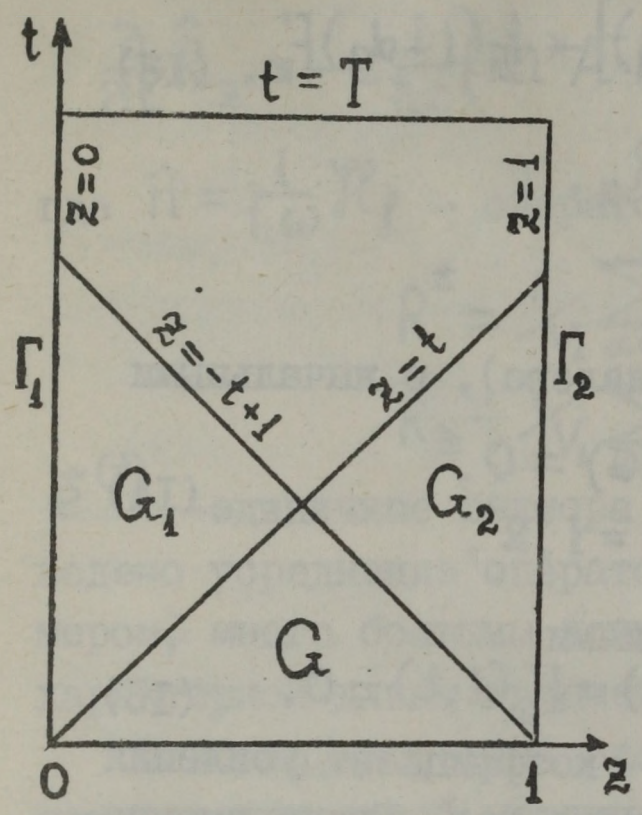


Рис. I.

Структура области П.

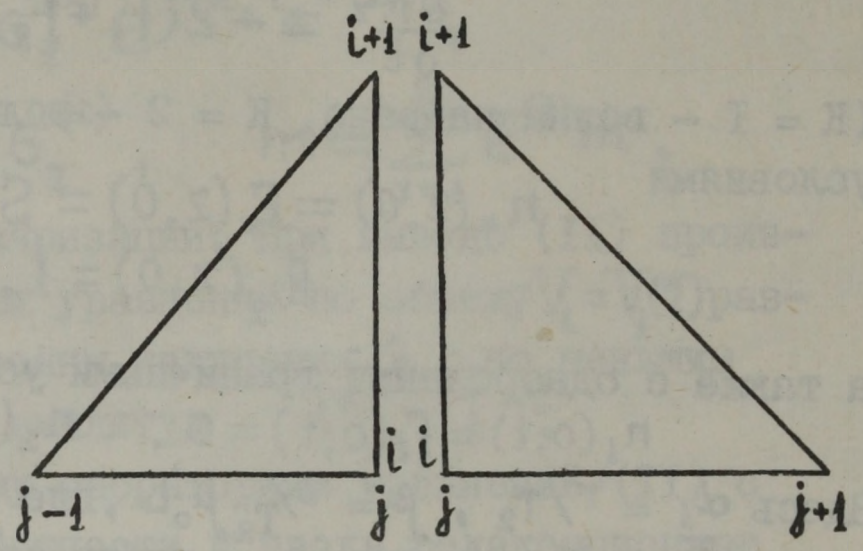


Рис. 2.

Шаблоны, используемые в разностной схеме.

Поставленная задача (I3)-(I5) удовлетворяет условиям корректности граничных условий и сопряжения начальных и граничных условий [3,4] .

Существование и ограниченность решения в области П можно доказать с помощью исследования мажорантной системы для (I3)-(I5) [3,4] . Такая система, полученная на основе продолженной системы для (I3), будет иметь вид [4]

$$\frac{\partial \mathcal{P}}{\partial t} = \mathcal{F}_1 \mathcal{P},$$

$$\frac{\partial v}{\partial t} = F_0(v) + F_0^{(1)}(v)\mathcal{P}$$

с начальными условиями

$$\mathcal{P}(0) = 0, \quad v(0) = v_0 .$$

Откуда $\mathcal{P}(t) \equiv 0$, а для v получим уравнение

$$\partial/\partial t\, v = a v^2 + b v ,$$

так что решение имеет вид

$$v(t) = \frac{b v_0}{(b + a v_0) e^{-bt} - a v_0} . \qquad (I8)$$

Здесь a и b представляют собой максимальные значения коэффициентов при линейных и квадратично зависящих от неизвестных функций членах системы (I3). Учитывая интересующие нас при проведении численных экспериментов значения параметров задачи и используя (I8), получим оценку снизу для T

$$T \geq \frac{1}{2} \ln(1 + 4/\beta). \qquad (I9)$$

Таким образом, в области П вида (I6) со значением T, удовлетворяющим (I9), гарантированы существование и единственность классического решения задачи (I3)-(I5) [5].

4. Алгоритм численного решения. Обратимся к численным методам решения поставленной задачи. Использовались два метода характеристик численного интегрирования: на основе модифицированного метода Эйлера и - на основе метода трапеций-, что приводит соответственно к явной и неявной разностным схемам.

Для написания схем введем сеточную область $\Pi_{h\bar{\tau}} = \{(z_j) \times (t_i)\}$, где $z_j = jh$, $j = 1, \ldots, K$, $t_i = i\bar{\tau}$, $i = 1, \ldots, M$, $h = 1/K$, $z_0 = 0$, $z_K = 1$, $t_0 = 0$, $t_M = T$. При этом будем использовать четырехточечный шаблон, представленный на рис. 2. Отметим, что левая половина шаблона-1 соответствует волне, бегущей направо, правая половина-2 - волне, бегущей налево; боковые стороны треугольников представляют собой отрезки характеристик семейства (I7), если $\bar{\tau} = h$, что в дальнейшем и будем предполагать.

В случае аналога метода Эйлера для первого, например, уравнения системы (I3) получим, обозначая $u(z_j, t_i) = u_j^i$, ($u = \{n_1, n_2, F_1, F_2, S, R_2\}$ - неизвестная вектор-функция),

$$u_{1j}^{i+1} - u_{1j-1}^{i} = h f_{1j-1}^{i}, \qquad (20)$$

где f -вектор правых частей (I3). Отсюда находим неизвестные функции на $i+1$ временном слое. Это соответствует обыкновенному дифференциальному уравнению вдоль отрезка характеристики $[(z_{j-1}, t_i), (z_j, t_{i+1})]$. Порядок аппроксимации можно повысить до $O(h^2)$, если применить модификацию

$$u_{1j}^{i+1} = u_{j-1}^{i} + \frac{h}{2}\left(f_{1j-1}^{i} + \tilde{f}_{1j}^{i+1}\right), \qquad (2I)$$

где $\tilde{f}$ вычислено с использованием значений неизвестных функций на i -ом слое, полученных по формулам (20). Для остальных уравнений (I3) соотношения (20)-(2I) записываются

также вдоль отрезков соответствующих характеристик.

В случае аналога метода трапеций получим следующую неявную схему

$$u_{1j}^{i+1} - u_{1j-1}^{i} = \frac{h}{2}\left(f_{1j-1}^{i} + f_{1j}^{i+1} \right). \qquad (22)$$

Разностное уравнение (22) аппроксимирует (I3) с точностью $O(\quad)$. Для решения нелинейных разностных уравнений применим итерационный метод

$$\left(u_{1j}^{i+1}\right)_{m+1} = u_{1j-1}^{i} + \frac{h}{2}\left(f_{1j-1}^{i} + \left(f_{1j}^{i+1}\right)_{m} \right), \qquad (23)$$

где m -номер итерации.

Справедливы следующие утверждения [4,5] :

I. Решение системы нелинейных разностных уравнений существует и ограниченно, итерационный процесс (23) сходится при $m \to +\infty$.

2. Построенное решение разностной задачи сходится к классическому решению задачи (I3)-(I5) со скоростью $O(h^2)$.

Следует отметить, проводя сравнение предложенных алгоритмов численного решения, что второй обладает большей вычислительной устойчивостью. Решение, полученное по формулам (20)-(2I) становится неустойчивым к погрешностям вычислений для достаточно больших значений t . Для малых t проведено сопоставление полученных по формулам (20)-(2I) и (22) результатов, давших хорошее совпадение. Для контроля проводился пересчет с вдвое меньшим шагом, а также качественное сравнение с решениями, полученными для более простых моделей процесса.

5. О с н о в н ы е р е з у л ь т а т ы. На рис. 3-5 представлены результаты численного интегрирования системы уравнений (I3). Рис. 3 дает представление о пространственно-временном развитии импульса СИ($\tau/T_2 = 10^{-2}$, $\mu_0 L = 100$). Из рисунка видно, что как на переднем фронте импульса (кривая I), так и на спаде импульса (кривая 3) происходит экспоненциальное нарастание интенсивности волны с увеличением z (для волны, бегущей вправо). Таким образом, две встречные волны в режиме СИ взаимодействуют очень слабо. Это обстоятельство находит свое отражение и в пространственной зависимости разности населенностей атомов R_z (рис. 5а, сплошные кривые). Видно, что разность населенностей очень мало меняется в пределах всего образца. Такой характер развития лавины СИ свидетельствует в

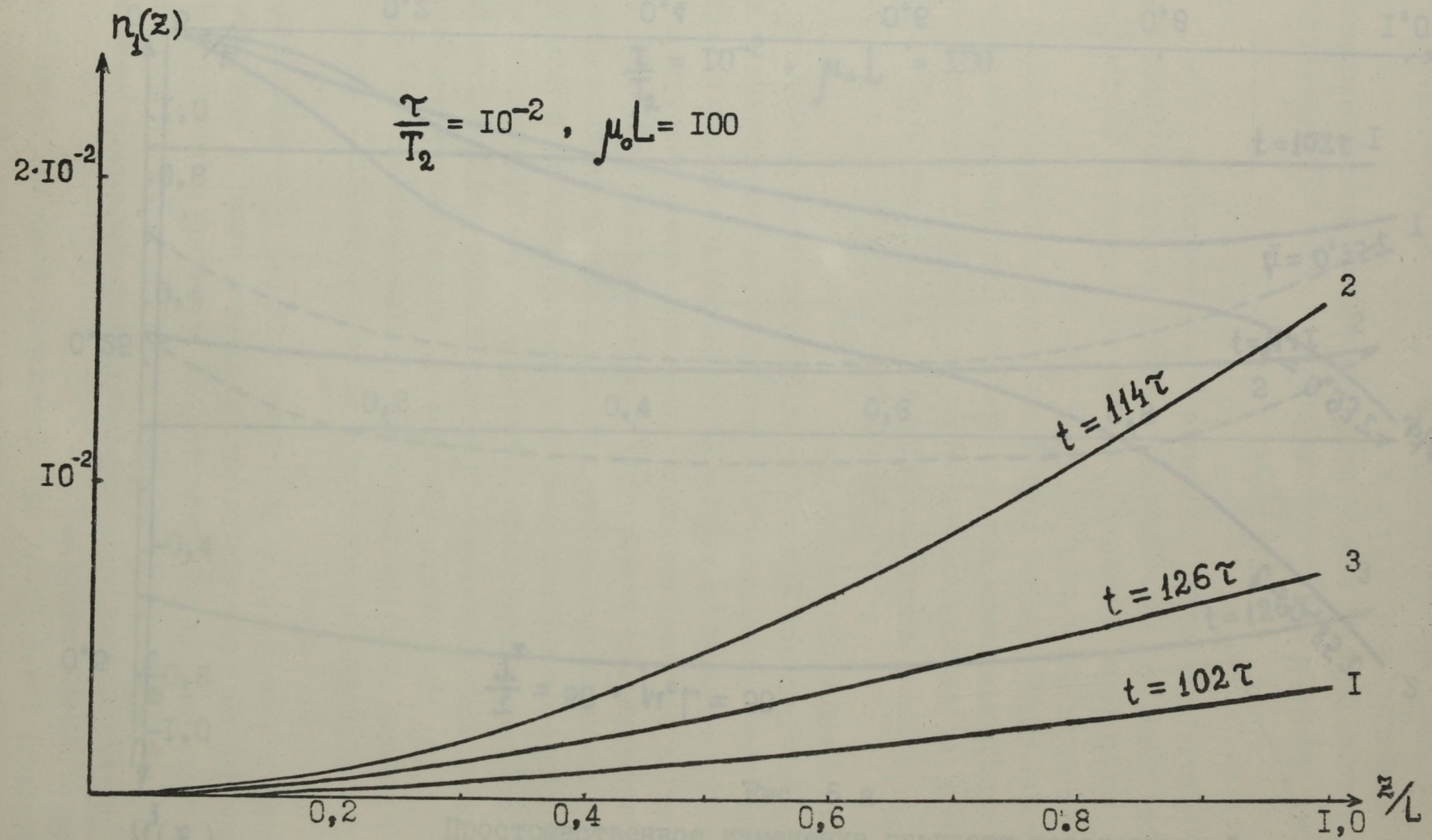


Рис. 3

Пространственное развитие сверхизлучения

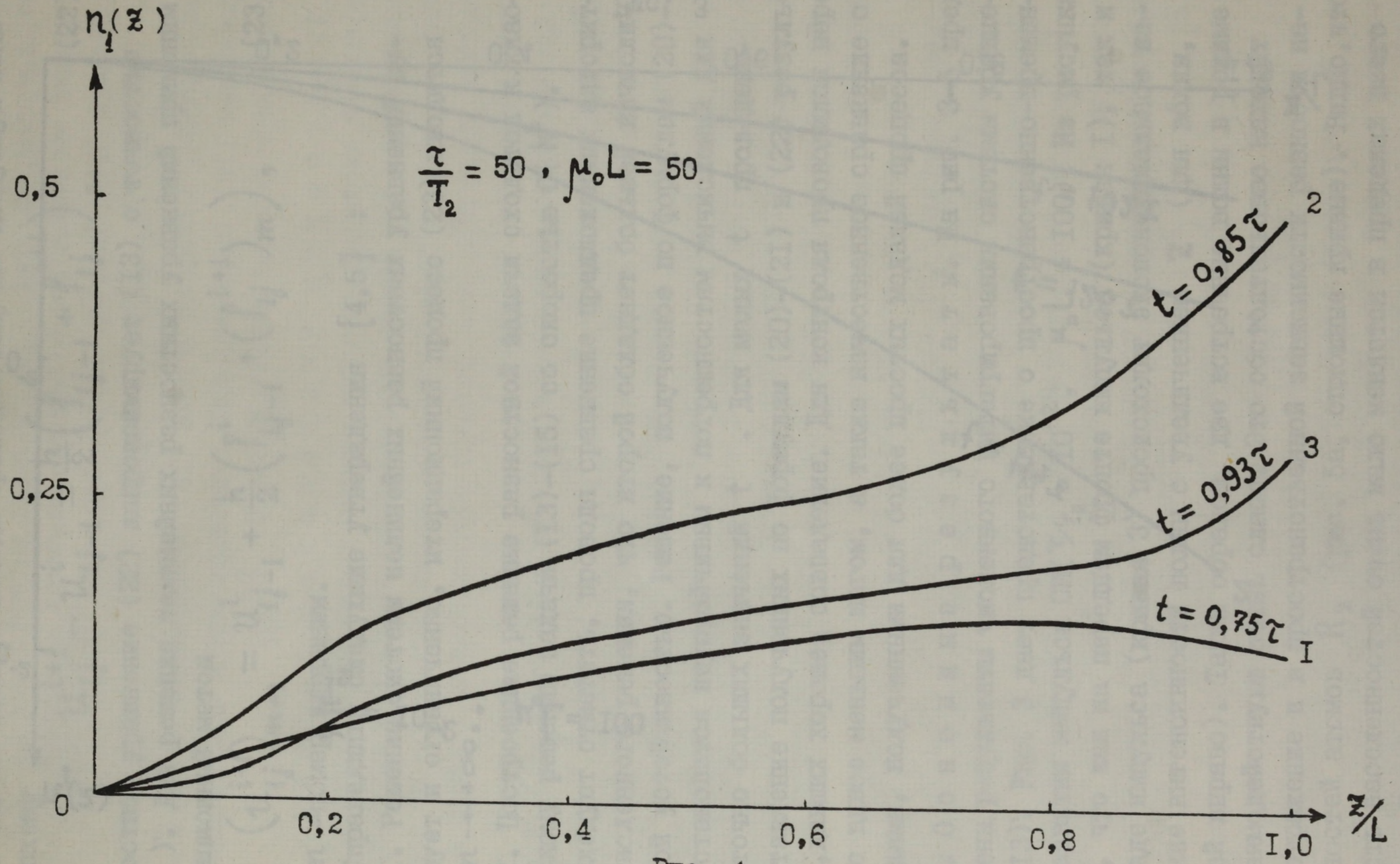


Рис. 4

Пространственное развитие суперлюминесценции

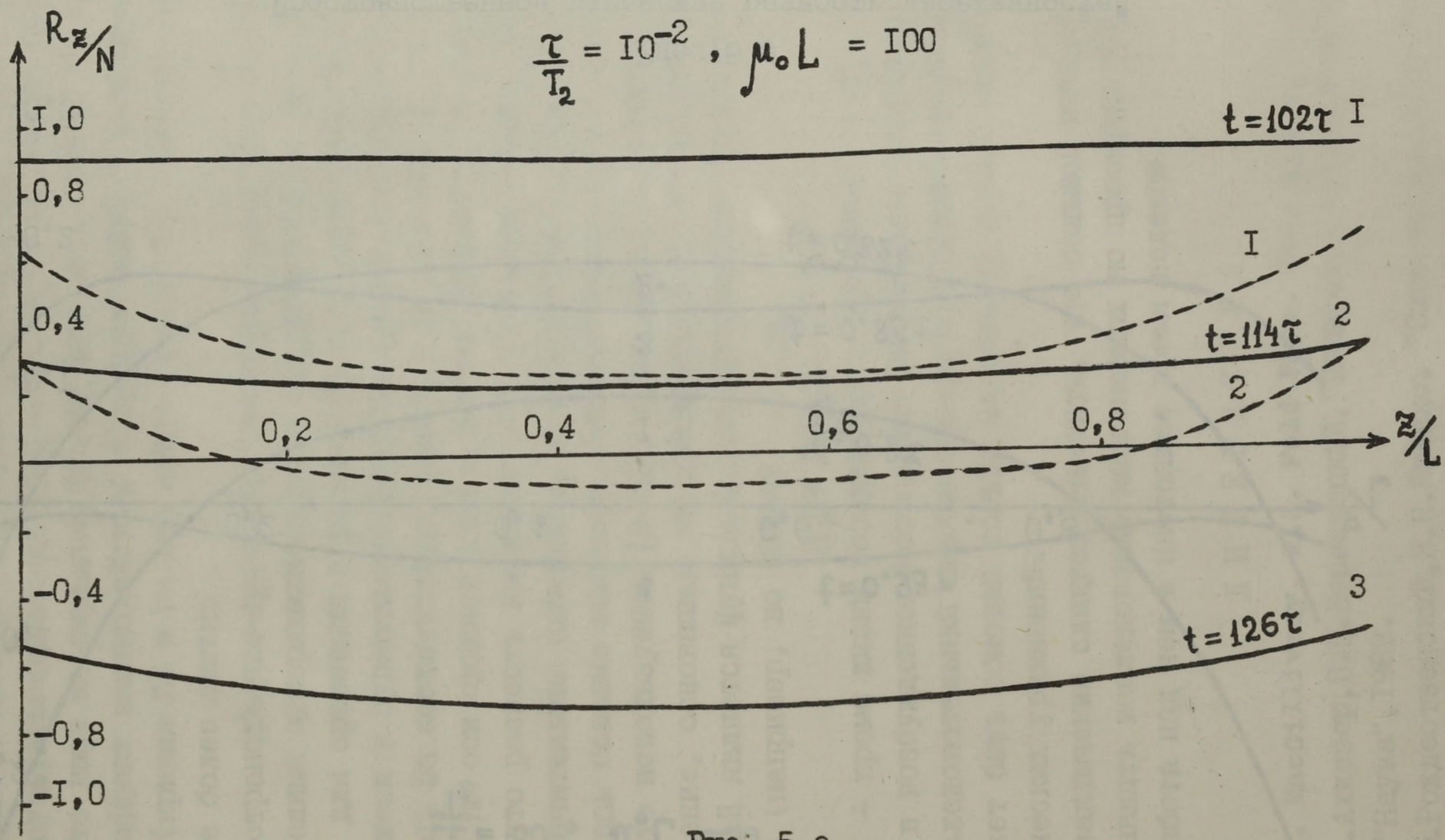


Рис. 5 а

Пространственное изменение разности населенностей
(режим сверхизлучения)

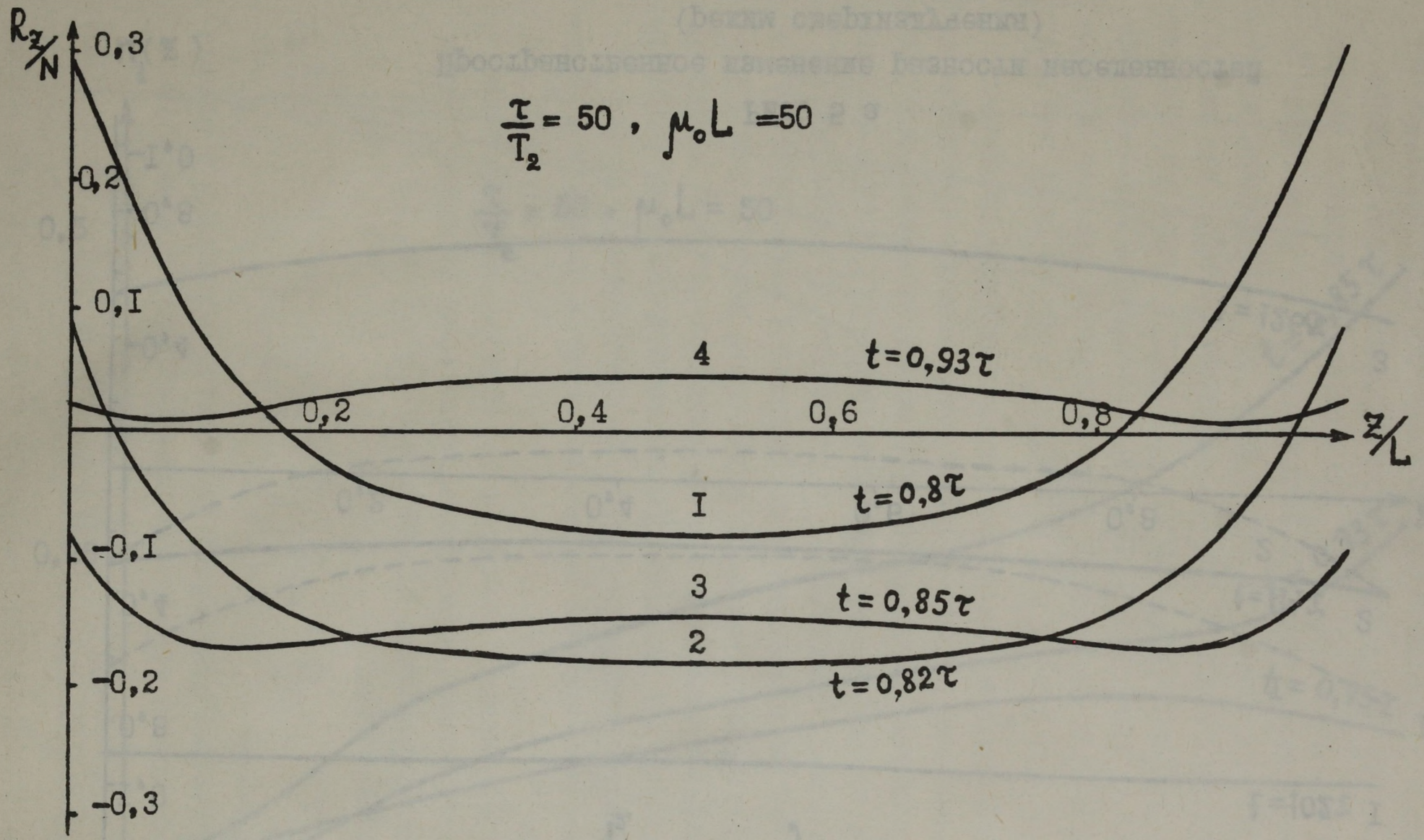


Рис. 5 б

Пространственное изменение разности населенностей
(режим суперлюминесценция)

пользу справедливости одномодовой теории СИ, которую мы использовали ранее [5-8] для анализа возможности наблюдения эффекта в гамма-диапазоне.

На рис. 4 показано пространственно-временное развитие импульса суперлюминесценции ($\tau/T_2 = 50$, $\int_0 L = 50$). Из рисунка видно, что даже на переднем фронте импульса (кривая I) проявляется эффект взаимодействия встречных волн. В максимуме импульса (кривая 2) и на его спаде (кривая 3) взаимное влияние волн еще более сильно .

Пространственно-временное поведение разности населенностей в режиме суперлюминесценции представлено на рис. 5б. На рис. 5а для сравнения с режимом СИ пунктиром нанесены кривые, относящиеся к суперлюминесцентному режиму (пунктирная кривая 2 на рис. 5а соответствует кривой I на рис. 5б, на рис. 5б масштаб по оси ординат взят в 4 раза крупнее). Из рис. 5б мы видим, что разность населенностей на концах образца претерпевает значительные изменения, а следовательно, $\partial/\partial t\, R_z$ принимает здесь большие значения. Последнее свидетельствует о том, что часто используемое [9-II] в теории суперлюминесценции приближение, основанное на предположении, что разность населенностей является функцией, экспоненциально затухающей во времени (например, по закону

$$R_z = 2N_B e^{-t/T_1} - N,$$

где T_1 - время жизни возбужденного состояния, а N_B - число атомов в возбужденном состоянии) не является оправданным.

Последовательный анализ кинетики суперлюминесцентных систем может быть основан только на решении полной системы квазиклассических уравнений.

В заключение следует отметить, что большой интерес может представлять исследование многомерных по пространству моделей СИ , работа над чем в настоящее время ведется.

Л И Т Е Р А Т У Р А

I. J.C. MacGillivray, M.S. Feld Phys. Rev, A14, II69, I976.
2. А.И.Ахиезер, В.Б.Берестецкий, "Квантовая электродинамика", М., "Наука", I969.
3. Б.Л.Рождественский, Н.Н.Яненко, "Системы квазилинейных уравнений", М., "Наука", I968.
4. О.Ю.Тихомиров, "Некоторые математические модели -излу-

чения в мессбауэровской резонансной среде", кандидат. диссерт. МГУ, 1978.

5. А.В.Андреев, В.Я.Галкин, Ю.А.Ильинский, О.Ю.Тихомиров, В сб. "Обработка и интерпретация физических экспериментов" Изд-во МГУ, стр.60, 1978.
6. А.В.Андреев, ЖЭТФ, 72, 1397, 1977.
7. А.Н.Тихонов, А.В.Андреев, В.Я.Галкин, О.Ю.Тихомиров, Труды межд.совещ.по прог. и прим. мат. методов реш.физ.задач. ОИЯИ, стр.9, 1978.
8. А.В.Андреев, Ю.А.Ильинский, Р.В.Хохлов, ЖЭТФ, 73, 1296, 1977.
9. Б.В.Чириков, ЖЭТФ, 44, 2017, 1963.
10. В.И.Воронцов, В.И.Высоцкий, ЖЭТФ, 66, 1528, 1974.
11. В.А.Бушуев, Р.Н.Кузьмин, О.Ю.Тихомиров, В сб."Обработка и интерпретация физических экспериментов" Изд-во МГУ, стр.91, 1976.

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССОВ УСИЛЕНИЯ И ГЕНЕРАЦИИ ИЗЛУЧЕНИЯ В ГАММА-ЛАЗЕРЕ

А.Н.Тихонов, В.А.Бушуев, В.Я.Галкин, Р.Н.Кузьмин, О.Ю.Тихомиров

Московский государственный университет им. М.В. Ломоносова, Москва

АННОТАЦИЯ

В настоящей работе ставятся и численно решаются некоторые математические задачи кинетики в гамма-лазере. В качестве основы для математического моделирования используется квазиклассический подход, при котором поле описывается классически: уравнениями Максвелла, а рабочая среда - квантовомеханически: уравнением Шредингера. Исследуются вопросы корректности поставленных задач, сходимости и устойчивости предложенных численных алгоритмов. На основе проведенных численных экспериментов делается ряд важных выводов о физике процесса.

ABSTRACT

Some mathematical problems of gamma-laser kinetic are considered and numerically solved. As a base for mathematical model a quasiclassical approach is used in which the laser field is described with the help of Maxwell's classical equations and working media by Schrödinger equation. Questions of right pouseness of considered problems are discussed. Besides questions of convergence and equalibration are considered. On the base of accomplished numerical experiments several important conclusions on the physical aspects of the process are made.

I. ВВЕДЕНИЕ

В настоящее время проявляется оольшой интерес к проблеме создания γ-лазера с длиной волны $\lambda \sim 0{,}1\text{-}1\,\text{Å}$ на основе мессбауэровских изотопов [1,2]. Гамма-лазер - это пока гипотетическое устройство, предназначенное для получения мощного остронаправленного монохроматического и когерентного излучения γ-квантов. Ожидается, что с появлением γ-лазеров произойдет не менее бурный рост исследований и приложений в науке и технике, чем в 60-е годы при создании оптических квантовых генераторов. В проблеме γ-лазера имеется ряд нерешенных вопросов теоретического плана. Требуют исследования также организация и проведение сложных, трудоемких и дорогостоящих экспериментов. Зачастую в решении подобных вопросов математическое моделирование и проектирование играет определяющую роль [3,4].

Исследование характера развития волны в γ-лазере приводит к постановке сложных математических задач, таких как интегрирование систем нелинейных дифференциальных уравнений в частных производных и т.п. При их решении возникает необходимость разработки численных алгоритмов и соответствующего программного обеспечения для ЭВМ. Уже первые публикации [5-9] показали, что временной характер кинетики γ-излучения может существенно изменить величину порога генерации, которая в стационарном случае имеет вид $K = \mu$. В этой связи уже на начальном этапе исследования возникает необходимость детального изучения кинетики в γ-лазере.

Цель настоящей работы заключается в исследовании с помощью математических моделей кинетики усиления и генерации излучения в γ-лазере, в разработке и уточнении существующих моделей явления на основе квазиклассического подхода, а также

в проведении численных экспериментов. Анализируются вопросы корректности математических постановок задач, численной реализации предложенных алгоритмов и физической интерпретации полученных результатов.

2. ПОСТАНОВКА СМЕШАННОЙ НАЧАЛЬНОЙ И ГРАНИЧНОЙ ЗАДАЧИ В РАМКАХ КВАЗИКЛАССИЧЕСКОГО ПОДХОДА

Приведенные в публикациях [5-7] постановки задач кинетики соответствуют некоторым частным случаям общей квазиклассической системы уравнений однопроходного двухуровневого γ-усилителя для медленно меняющейся амплитуды $A(x,t)$ вектор-потенциала электромагнитного поля. В квазиклассическом приближении, введя безразмерные переменные $x = x/\ell$ и $t = t/\tau$, система уравнений γ-усилителя, полученная на основании уравнений Максвелла и Шредингера [10], может быть представлена в следующем виде

$$\mu_1 \frac{\partial A}{\partial t} + \frac{\partial A}{\partial x} = p - \frac{1}{2}A, \qquad (1)$$

$$\frac{\partial p}{\partial t} = C\Delta n A - (i\varepsilon + \frac{\Gamma\tau}{2})p, \qquad (2)$$

$$\frac{\partial \Delta n}{\partial t} = -D(Ap^* + pA^*) - (1+\Delta n), \qquad (3)$$

где $\mu_1 = \ell/\tau c$, $\ell = 1/\mu$ – длина пробега γ-квантов, μ – коэффициент нерезонансного поглощения, τ – время жизни возбужденного состояния, c – скорость света, $\varepsilon = \tau\cdot(\omega - \omega_0)$, ω – частота γ-излучения, $\hbar\omega_0$ – энергия ядерного γ-перехода,

$$C = (K\ell)(\Gamma\tau)/4,\ K = \sigma_0 n_0,\ D = 2\tau/\hbar n_0 \ell\lambda, \qquad (4)$$

$$\sigma_0 = \frac{\lambda^2}{2\pi}\frac{f}{1+\alpha}\frac{1}{\Gamma\tau}\frac{g_2}{g_1};$$

здесь $K\cdot\Delta n$ – коэффициент резонансного усиления в максимуме при $\omega = \omega_0$, $n_0 = n_1 + n_2$ – плотность рабочих ядер, $\Delta n(x,t) = (g_{12}n_2 - n_1)/n_0$ – нормированная к n_0 плотность инверсной населенности, $g_{12} = g_1/g_2$, g_1 и g_2 – статистические веса нижнего состояния I и верхнего

2; f – вероятность безотдачного поглощения Лэмба–Мессбауэра, α – полный коэффициент внутренней электронной конверсии, Γ – ширина линии излучения, $p(x,t)\cdot(c/\lambda l)$ – ток ядерного перехода.

При таком подходе электромагнитное поле описывается квазиклассически, мессбауэровская резонансная среда – квантовомеханически. Следует отметить, что приведенные уравнения квазиклассического приближения – неединственный способ описания кинетики: в работах [II–I4] рассмотрены математические модели, полученные на основе последовательного квантового подхода.

Введя векторные обозначения, перепишем (I)–(3) в виде

$$u_t(x,t) + \hat{G}\cdot u_x(x,t) = F(u,x,t), \qquad (5)$$

где $u = (A, \hat{p}, \Delta n)^T$ – трехмерный вектор–столбец неизвестных функций, $\hat{G}$ – матрица 3 x 3, $G_{11} = \mu_1^{-1}$, $G_{ij} = 0$ при $i+j > 2$, $F = (\mu_1^{-1}(p-0{,}5A), CA\Delta n - (i\varepsilon + \frac{\Gamma\tau}{2})p, -D(Ap^* + pA^*) - (1+\Delta n))^T$ – вектор правых частей системы (5).

Из вида системы дифференциальных уравнений в частных производных I–го порядка (5) можно заключить, что она является полулинейной [I5,I6], т.е. разновидностью квазилинейных систем, у которых дифференциальный оператор не зависит от неизвестной функции u, а нелинейность есть только в правой части. Все собственные значения (5) действительны и равны соответственно $\xi_1 = \mu_1^{-1}$, $\xi_2 = \xi_3 = 0$, откуда следует гиперболичность системы в широком смысле. Уравнения для опорных характеристик определяются как $x = \mu_1^{-1} t$, $x = 0$, причем последняя характеристика кратная.

Для системы (5) в приложениях к задачам кинетики представляет интерес постановка смешанной начальной и граничной задач. Решение надлежат получить в области $\{[0,X]\times[0,T]\}$ для положительных значений переменных x, t. Начальные и граничные условия задаются соответственно на положительных полуосях x и t. В работе [I7] сформулированы достаточные условия существования и ограниченности решения смешанной начальной и граничной задачи кинетики.

Предположим, что в начальный момент времени среда инвертирована и извне на нее падает электромагнитная волна. Такой постановке задачи соответствуют следующие начальные условия

$$A(x,0)=0,\ p(x,0)=0,\ \Delta n(x,0)=(1+g_{12})\eta_0-1, \quad (6)$$

где $\eta_0 = n_2(0)/n_0$ – начальная нормированная концентрация возбужденных ядер. Граничный режим определяется формой входного сигнала

$$A(0,t)=\varphi(t),\ \varphi(t)\in C[0,T]. \quad (7)$$

Заданное в таком виде граничное условие (7) отвечает условиям теоремы о корректности граничных условий [I8] . В случае невыполнения условия согласования $\varphi(0)=0$, по характеристике $x=\mu_1^{-1}t$ будет распространяться разрыв решения для $A(x,t)$. Функции p и Δn будут оставаться непрерывными. Можно показать, что в области

$$T \le \frac{1}{a}\ln\left(1+\frac{a}{v_0 b}\right)$$

существует единственное классическое решение задачи (5)–(7). Здесь a и b – максимальные значения коэффициентов при квадратичных и линейных слагаемых правых частей (5) соответственно ; v_0 – максимум нормы начальных и граничных условий.

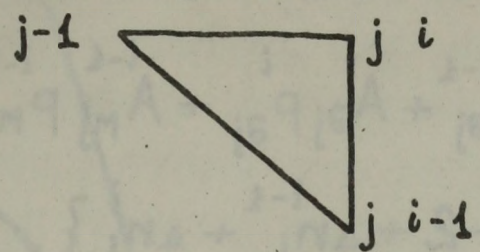
Поскольку даже для короткоживущих изомеров $l \ll c\tau$, то параметр μ_1 является малым. Как нетрудно видеть, отбрасывание слагаемого $\mu_1 (\partial A/\partial t)$ эквивалентно переходу к новым переменным $\eta = t - x\mu_1$, $\xi = x$. В результате такой замены мы придем к локальному времени η , отсчитываемому в точке x от момента прихода γ-волны в эту точку.

3. ЧИСЛЕННОЕ РЕШЕНИЕ ЗАДАЧИ ОБ УСИЛЕНИИ ВНЕШНЕГО СИГНАЛА

Для численного интегрирования системы (5)–(7) применены два варианта метода характеристик [I5,I6] : многомерные аналоги метода Эйлера и метода трапеций. Для написания разностной схемы введем сеточную область $\Pi_{h\bar\tau}$

$$\Pi_{h\bar\tau}=\left\{\begin{array}{l}(x_j)\times(t_i);\ x_j=jh;\ t_i=i\bar\tau;\ j=1,2,\dots,M;\\ i=1,2,\dots,k;\ x_0=0;\ x_M=X;\ t_0=0,\ t_k=T\end{array}\right\}. \quad (8)$$

При этом будем использовать трехточечный шаблон вида (9), где катеты – отрезки характеристик



(9)

Разностная аппроксимация системы (5) в случае аналога модифицированного метода Эйлера имеет вид

$$\frac{A^i_{эj} - A^i_{эj-1}}{h} = P^i_{эj-1} - \frac{1}{2} A^i_{эj-1},$$

$$\frac{A^i_{мj} - A^i_{мj-1}}{h} = P^i_{мj-1} - \frac{1}{2} A^i_{мj-1},$$

$$\frac{P^i_{эj} - P^{i-1}_{эj}}{\bar{\tau}} = C\Delta n^{i-1}_j A^{i-1}_{эj} - \frac{\Gamma\tau}{2} P^{i-1}_{эj} + \varepsilon P^{i-1}_{мj},$$

$$\frac{P^i_{мj} - P^{i-1}_{мj}}{\bar{\tau}} = C\Delta n^{i-1}_j A^{i-1}_{мj} - \frac{\Gamma\tau}{2} P^{i-1}_{мj} - \varepsilon P^{i-1}_{эj}, \qquad (10)$$

$$\frac{\Delta n^i_j - \Delta n^{i-1}_j}{\bar{\tau}} = -D\left(A^{i-1}_{эj} P^{i-1}_{эj} + A^{i-1}_{мj} P^{i-1}_{мj}\right) - \left(1 + \Delta n^{i-1}_j\right);$$

где $A^i_{эj}$, $A^i_{мj}$ и $P^i_{эj}$, $P^i_{мj}$ – действительные и мнимые части $A(x_j, t_i)$ и $P(x_j, t_i)$ соответственно, $\Delta n^i_j = \Delta n(x_j, t_i)$. Схема (10) первого порядка точности, явная, поэтому можно выразить значения неизвестных функций в точке сетки (x_j, t_i) и произвести коррекцию (10), используя найденные значения. При этом правые части (10) заменяются на новые с учетом коррекции. Например, для первого уравнения имеем

$$\frac{A^i_{эj} - A^i_{эj-1}}{h} = \frac{1}{2}\left(P^i_{эj-1} - \frac{1}{2} A^i_{эj-1} + \tilde{P}^i_{эj} - \frac{1}{2}\tilde{A}^i_{эj-1}\right), \qquad (11)$$

где $\tilde{A}$ и $\tilde{P}$ вычислены согласно (10). Такая коррекция позволяет повысить порядок аппроксимации до $0(h^2 + \bar{\tau}^2)$. Разностный алгоритм численного интегрирования (10)–(11) прост в применении, однако остается устойчивым к погрешностям вычисления для сравнительно малых значений t.

Наряду с описанной рассмотрена разностная схема, соответ-

ствующая многомерному аналогу метода трапеций. Например, для уравнения (3) эта схема примет вид

$$\frac{\Delta n_j^i - \Delta n_j^{i-1}}{\bar{\tau}} = -\frac{1}{2}\{D(A_{\partial j}^{i-1} P_{\partial j}^{i-1} + A_{\partial j}^{i} P_{\partial j}^{i} + A_{Mj}^{i-1} P_{Mj}^{i-1} + A_{Mj}^{i} P_{Mj}^{i}) + 2 + \Delta n_j^{i-1} + \Delta n_j^{i}\}. \quad (12)$$

Нелинейные разностные уравнения типа (12) аппроксимируют (5) с точностью $O(h^2 + \bar{\tau}^2)$. Данная схема является неявной и здесь могут быть применены различные итерационные методы. Положив $\bar{\tau} = h$, запишем вид выбранного итерационного метода для (12)

$$[\Delta n_j^i]_{m+1} = \Delta n_j^{i-1} - \frac{h}{2}\{D(A_{\partial j}^{i-1} P_{\partial j}^{i-1} + [A_{\partial j}^{i}]_m [P_{\partial j}^{i}]_m + A_{Mj}^{i-1} P_{Mj}^{i-1} + [A_{Mj}^{i}]_m [P_{Mj}^{i}]_m + 2 + \Delta n_j^{i-1} + [\Delta n_j^{i}]_m\}, \quad (13)$$

где m – номер итерации. Начальные приближения возьмем нулевыми. Для предложенного метода (12)–(13) справедливы следующие утверждения [18]: 1° Решение нелинейной разностной системы уравнений вида (12) существует и остается ограниченным при выборе достаточно малого значения h. Итерационный процесс (13) сходится. 2° Построенное решение разностной задачи сходится к классическому решению (5)–(7) со скоростью $O(h^2)$.

Укажем характерные значения параметров, используемые в расчетах: $\lambda \sim 0{,}1\text{–}1$ Å, $n_0 \sim 10^{21}\text{–}5\cdot10^{22}$ см$^{-3}$, $\tau \sim 10^{-7}\text{–}10^{3}$ сек, $f \sim 0{,}1\text{–}1$, $\alpha \sim 1\text{–}100$, $l \sim 10^{-2}\text{–}1$ см, $g_{12} \sim 1$, $\varepsilon = 0$. На рис. 1–4 приведены результаты расчетов для временного и пространственного развития γ-волны усиленного постоянного внешнего сигнала $\varphi(t) = A_0 = 10^{-11}$ г$^{1/2}$ см$^{1/2}$/сек. Это значение соответствует амплитуде излучения мессбауэровского источника активностью примерно 1 кюри.

Кривые временного развития $A(t)$ при $C = 10$ и $\eta_0 = 1$ на рис. 1 демонстрируют важность учета характера кинетики и ее зависимость от ширины линии. Действительно, при переходе от случая с $x = l$ к случаю $x = 2l$ при $\Gamma\tau = 10$ максимальная амплитуда возрастает в 1,4 раза, в то время как в рамках стационарной модели возрастание должно произойти в $\exp[(K-\mu)l/2] = \exp(3/2) \approx 4{,}5$ раза (напомним, что $K/\mu = 4C/\Gamma\tau$,

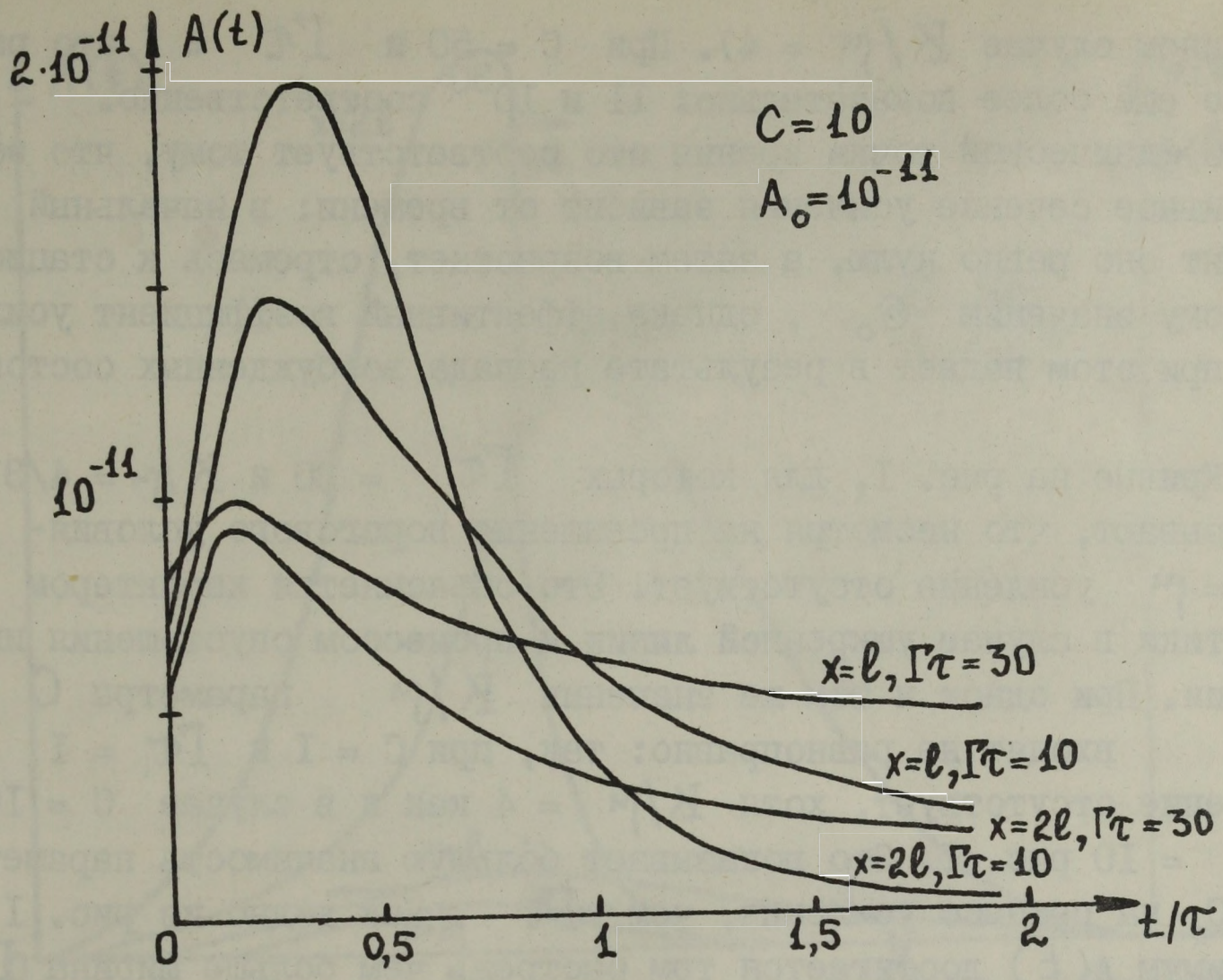


Рис. I. Зависимость кинетики усиления от ширины линии Γ .

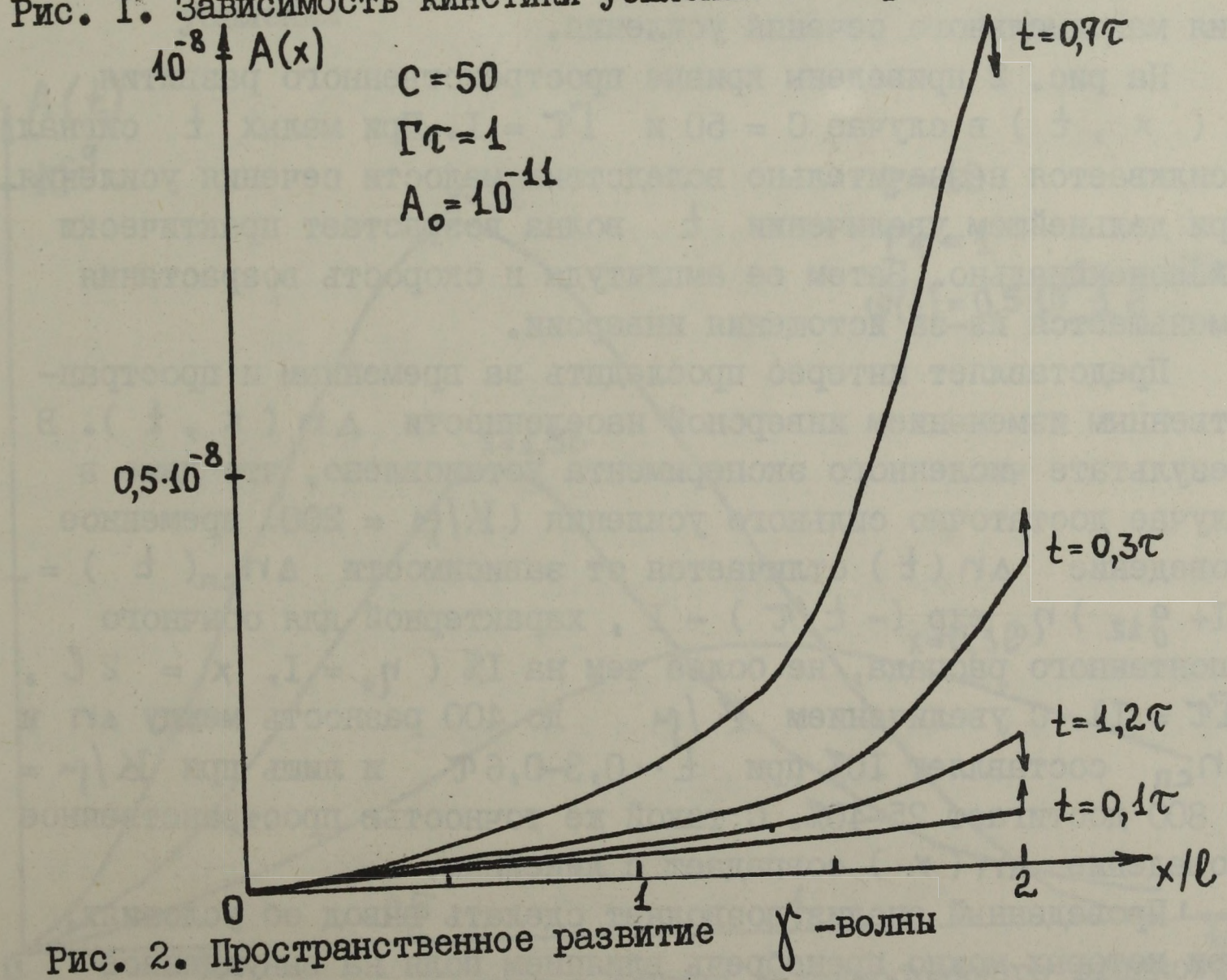


Рис. 2. Пространственное развитие $\int$ -волны

в данном случае $K/\mu = 4$). При $C = 50$ и $\Gamma\tau = 1$ это различие ещё более показательно: 11 и 10^{30} соответственно.

С физической точки зрения это соответствует тому, что эффективное сечение усиления зависит от времени: в начальный момент оно равно нулю, а затем возрастает, стремясь к стационарному значению σ_0, однако эффективный коэффициент усиления при этом падает в результате распада возбужденных состояний.

Кривые на рис. 1, для которых $\Gamma\tau = 30$ и $K/\mu = 4/3$, показывают, что несмотря на превышение порогового условия $K = \mu$ усиление отсутствует. Это объясняется характером кинетики в случае уширенной линии и процессом опустошения инверсии. При одном и том же значении K/μ параметры C и $\Gamma\tau$ входят не равноправно: так, при $C = 1$ и $\Gamma\tau = 1$ усиление отсутствует, хотя $K/\mu = 4$ как и в случае $C = 10$, $\Gamma\tau = 10$ рис. 1. Это показывает бóльшую значимость параметра C на процесс усиления, чем $\Gamma\tau$. Как видно из рис. 1, максимум $A(t)$ достигается тем быстрее, чем больше ширина Γ, что вполне понятно с точки зрения временной эволюции достижения максимального сечения усиления.

На рис. 2 приведены кривые пространственного развития $A(x, t)$ в случае $C = 50$ и $\Gamma\tau = 1$. При малых t сигнал усиливается незначительно вследствии малости сечения усиления. При дальнейшем увеличении t волна возрастает практически экспоненциально. Затем ее амплитуда и скорость возрастания уменьшается из-за истощения инверсии.

Представляет интерес проследить за временным и пространственным изменением инверсной населенности $\Delta n(x, t)$. В результате численного эксперимента установлено, что даже в случае достаточно сильного усиления ($K/\mu = 200$) временное поведение $\Delta n(t)$ отличается от зависимости $\Delta n_{сп}(t) = (1 + g_{12})\eta_0 \exp(-t/\tau) - 1$, характерной для обычного спонтанного распада, не более чем на 1% ($\eta_0 = 1$, $x = 2l$, $\Gamma\tau = 1$). С увеличением K/μ до 400 разность между Δn и $\Delta n_{сп}$ составляет 10% при $t \sim 0{,}3\text{–}0{,}6\,\tau$ и лишь при $K/\mu = 800$ достигает 25–40%. С такой же точностью пространственное поведение $\Delta n(x)$ совпадает с линейным.

Проведенный анализ позволяет сделать вывод об условиях, при которых можно пренебречь влиянием поля на вынужденное

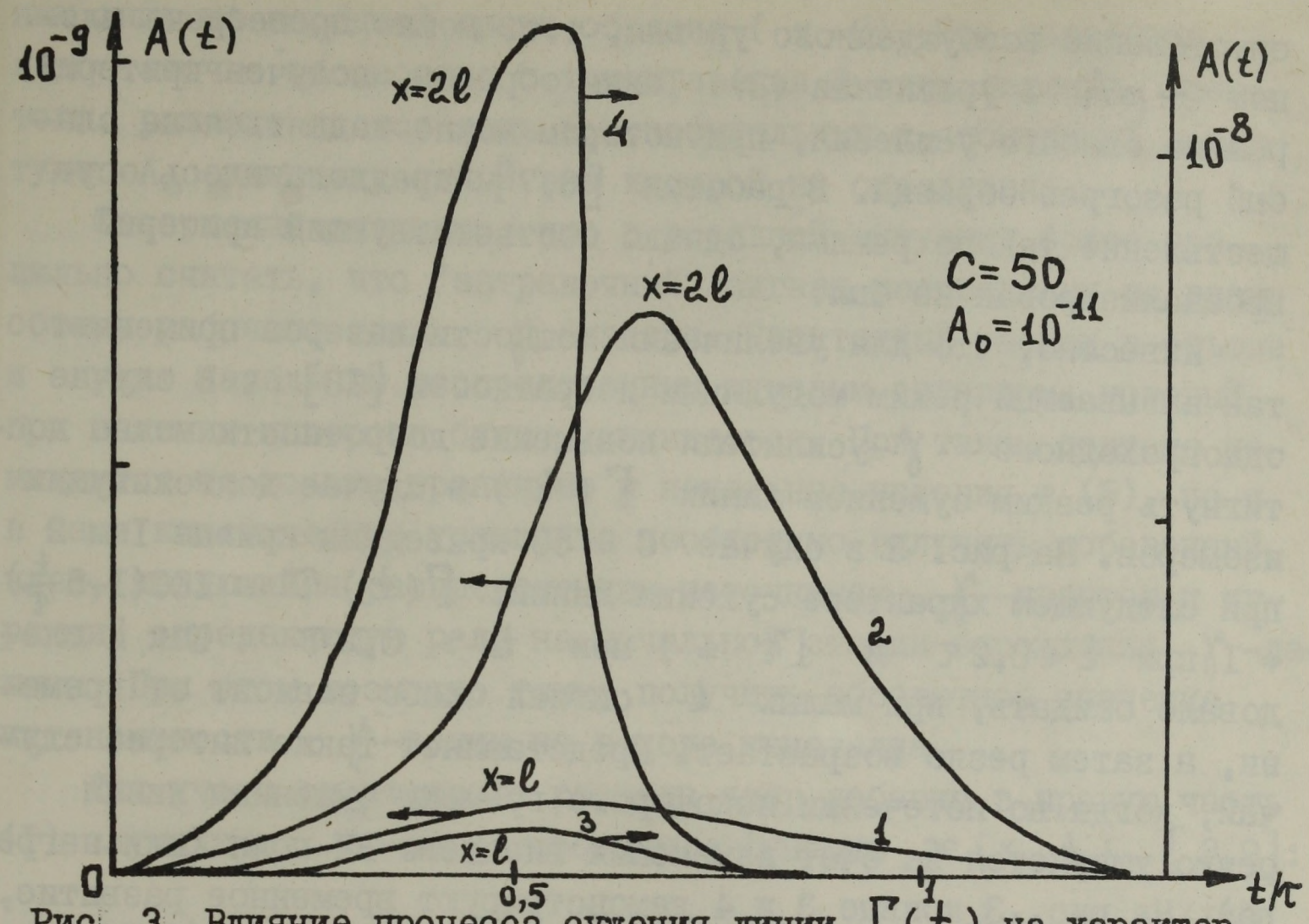


Рис. 3. Влияние процесса сужения линии $\Gamma(t)$ на характер усиления

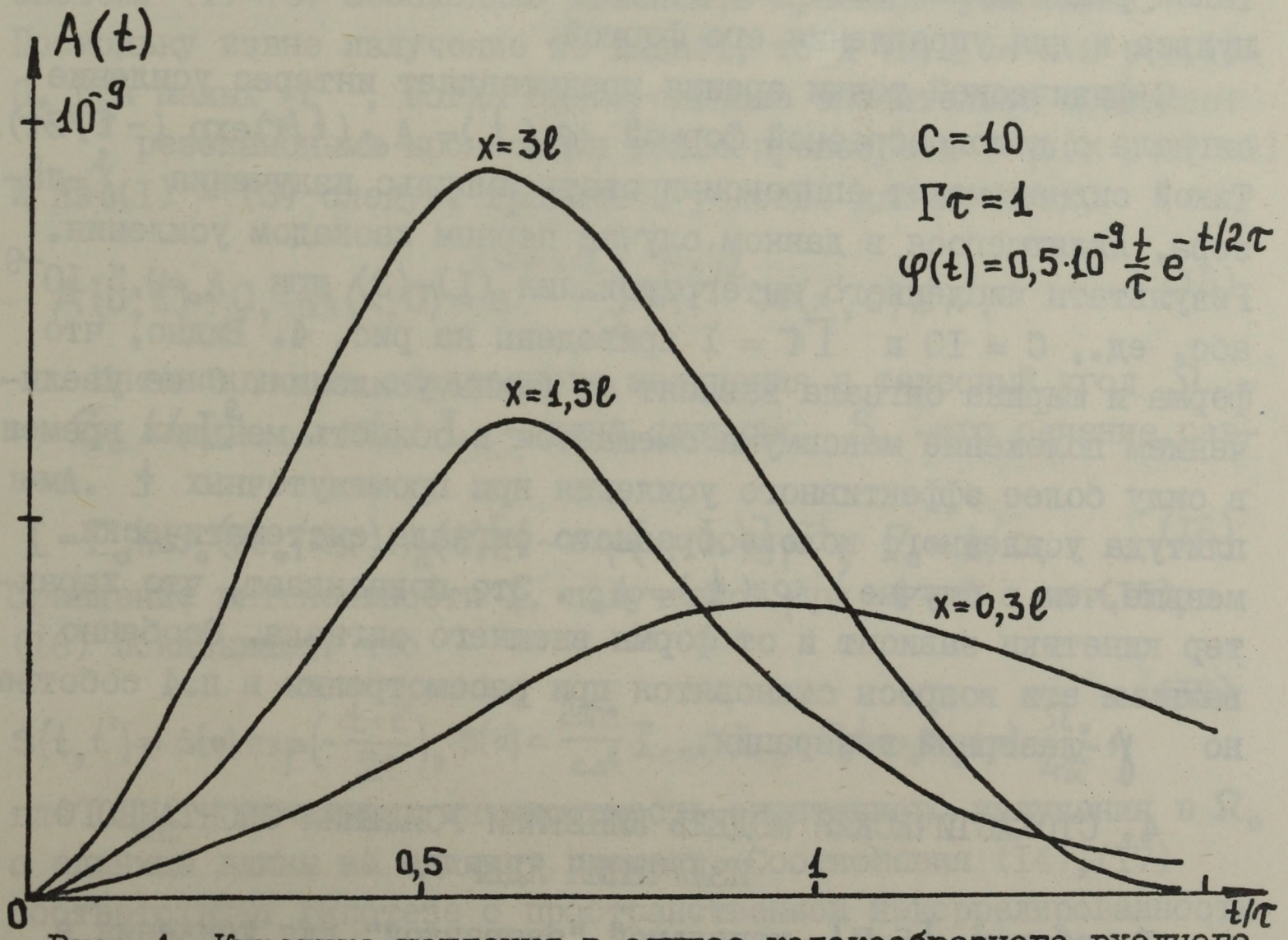


Рис. 4. Кинетика усиления в случае колокообразного входного сигнала

опустошение возбужденного уровня, т.е. можно пренебречь членом $\sim pA$ в уравнении (3). Таким образом, получен критерий режима слабого усиления, при котором можно надеяться на слабый разогрев образца. В работах [6,7] предполагалось осуществление такого режима, однако соответствующий критерий проанализирован не был.

Известно, что для увеличения мощности лазеров применяют так называемый режим модуляции добротности [10] . В случае однопроходного γ-усилителя повышение добротности можно достигнуть резким сужением линии $\Gamma(t)$ в случае долгоживущих изомеров. На рис. 3 в случае $C = 50$ приведены кривые I и 2 при следующем характере сужения линии: $\Gamma(t)\cdot\tau = 100(1-5\frac{t}{\tau})+$ $+ 1$ при $t \leqslant 0{,}2\tau$ и $\Gamma\tau = 1$ при $t > 0{,}2\tau$. Как и следовало ожидать, при малых t сигнал слабо зависит от времени, а затем резко возрастает. Представляет также интерес случай, когда по истечении некоторого времени суженная линия резко уширяется за счет изменения внешнего РЧ-поля (или нагрева). На рис. 3 кривые 3 и 4 демонстрируют временное развитие, когда $\Gamma\tau = 1$ при $t \leqslant (2/3)\tau$ и $\Gamma\tau = 50$ при $t > (2/3)\tau$. Такой режим может быть полезен для целей обострения γ-импульса и для управления его формой.

С физической точки зрения представляет интерес усиление сигнала с колокообразной формой $\varphi(t) = A_0\cdot(t/\tau)\exp(-t/2\tau)$. Такой сигнал может аппроксимировать импульс излучения γ-лазера, являющегося в данном случае первым каскадом усиления. Результаты численного интегрирования (I)-(3) при $A_0 = 0{,}5\cdot 10^{-9}$ абс. ед., $C = 10$ и $\Gamma\tau = 1$ приведены на рис. 4. Видно, что форма и ширина сигнала зависят от длины усиления. С ее увеличением положение максимума смещается в область меньших времен в силу более эффективного усиления при промежуточных t .Амплитуда усиленного колокообразного сигнала систематически меньше, чем в случае $\varphi(t) = A_0$. Это показывает, что характер кинетики зависит и от формы внешнего сигнала. Особенно важными эти вопросы становятся при рассмотрении в п.4 собственно γ-лазерной генерации.

4. СТОХАСТИЧЕСКАЯ МОДЕЛЬ КИНЕТИКИ УСИЛЕНИЯ СПОНТАННОГО ИЗЛУЧЕНИЯ ЯДЕР

В работах [5-7] начальной "затравкой" для усиления в γ-лазере выбрано поле спонтанного распада ядер, сосредоточен-

ных лишь на передней грани образца ($x=0$). Это положение недостаточно обосновано с точки зрения физики явления, кроме того, решение дается лишь в относительных к спонтанной амплитуде A_0 единицах, явный вид которой не определен.

Совершенно очевидно, что в реальной ситуации более правильно считать, что "затравочный" сигнал распределен по всему объему рабочего тела γ-лазера. Спонтанный распад в объеме и служит начальным распределенным шумовым сигналом, который при наличии инверсии будет усиливаться. При таком подходе изменяются не только граничные и начальные условия в (5), но и в сами кинетические уравнения необходимо включить добавочный член, являющийся распределенным источником γ-квантов и играющий определяющую роль на начальной стадии зарождения γ-лавины. При этом возможно также получить абсолютное значение интенсивности γ-волны на выходе кристалла.

Для учета спонтанного распада ядер добавим в правую часть (I) случайную δ-коррелированную функцию $\varkappa(x,t)$ [8,9]:

$$\langle \varkappa(x,t)\rangle = 0,\ \langle \varkappa(x,t)\,\varkappa^*(x',t')\rangle = S(t,t')\ell^2\delta(x-x'). \qquad (I4)$$

Систему (I)–(3) необходимо дополнить краевыми условиями. Поскольку извне излучение не падает, то $A(0,t)=0$ и $p(0,t)=0$. При малых t, когда спектр сигнала значительно превышает Γ, резонансными процессами можно пренебречь ($p(x,0)=0$) и из (I) – (3) следует граничное условие для A. Итак,

$$A(0,t)=0,\quad A(x,0)=e^{-\mu x/2}\int_0^x e^{\mu x'/2}\,\varkappa(x',0)\,dx'. \qquad (I5)$$

Интенсивность спонтанного излучения в телесный угол $\Omega_0 = S_0/4L^2$, где L –длина стержня, S_0 –его сечение, равна

$$I=\Gamma_0\hbar\omega_0(\Omega_0/4\pi)n_2(t)[1-\exp(-\mu L)]\mu^{-1};\ \Gamma_0=1/\tau. \qquad (I6)$$

Сравнение интенсивности I, полученной при $t\ll\tau$ из (I5), с (I6) показывает, что

$$S(t,t')=S(0)\exp\left(-\frac{t+t'}{2\tau}\right),\ S(0)=\frac{2\pi c}{\omega^2}I_{сп},\ I_{сп}=\Gamma_0\hbar\omega_0 n_2(0)\frac{\Omega_0}{4\pi}, \qquad (I7)$$

где $I_{сп}$ – начальная интенсивность спонтанного излучения в Ω_0 с единицы длины на единицу площади. Соотношения (I4), (I7) соответствуют гипотезе о пространственной некоррелированности экспоненциально затухающих по времени спонтанных источников.

Строго говоря, систему (5) нужно дополнить учетом волны A_2, бегущей влево. Однако, при малом усилении можно пренебречь вынужденным истощением населенности n_2, при этом уравнения для A и A_2 расщепляются и решаются независимо.

Рассмотрим задачу в предположении заданного закона изменения $\Delta n(t) = n_0[g e^{-t/\tau} - 1]$, где $g = (1 + g_{12})\eta_0$. Это позволяет линеаризовать задачу и соответствует случаю слабого усиления (см.п.3). Из (1)–(3) легко получить уравнение для $A(x, t)$:

$$\frac{\partial}{\partial t}\left(\frac{\partial A}{\partial x} + \frac{\mu A}{2}\right) + \Omega\left(\frac{\partial A}{\partial x} + \frac{\mu}{2}A\right) - \tilde{C}(t)A = \frac{\partial \varkappa}{\partial t} + \Omega\varkappa, \qquad (18)$$

где $\tilde{C}(t) = C_1[g \cdot \exp(-t/\tau) - 1]$, $C_1 = C\mu/\tau$, $\Omega = i(\omega - \omega_0) + (\Gamma/2)$. Представляя $A = \bar{A} \cdot \exp\left[-\left(\frac{\mu x}{2} + \int_0^t \Omega(t')\,dt'\right)\right]$, получим для $\bar{A}$ гиперболическое уравнение 2-го порядка с данными на характеристиках [19]

$$\frac{\partial^2 \bar{A}}{\partial x \partial t} - \tilde{C}(t)\bar{A} = e^{\frac{\mu x}{2} + \int_0^t \Omega(t')dt'}\left(\frac{\partial \varkappa}{\partial t} + \Omega\varkappa\right) \qquad (19)$$

с начальными и граничными условиями, определяемыми из (15).

При решении неоднородного уравнения (19) применим метод функции Римана [19], которая в данном случае является решением уравнения

$$\frac{\partial^2 R}{\partial x \partial t} - \tilde{C}(t)R = 0, \qquad (20)$$

где $R = R(x, t; \xi, \eta)$ ищется в области $0 \le x \le \xi$, $0 \le t \le \eta$ и удовлетворяет граничным условиям $R(x, \eta; \xi, \eta) = 1$, $R(\xi, t; \xi, \eta) = 1$. Применяя замену переменных в (20)

$$t' = C_1[g\tau(1 - e^{-t/\tau}) - t],$$

получим уравнение, решение которого известно [19]. Возвращаясь к старым переменным, окончательно получим

$$R(x,t;\xi,\eta) = \begin{cases} I_0(z) \text{ при } \eta < \tau \ln g, \\ \left.\begin{array}{l} I_0(z) \text{ при } 0 \le t \le \Phi(\eta) \\ J_0(z) \text{ при } \Phi(\eta) \le t \le \eta \end{array}\right\} \tau \ln g \le \eta \le t_0, \\ J_0(z) \text{ при } \eta > t_0, \end{cases} \qquad (21)$$

где

$$z = 2\left|C_1(x-\xi)[g\tau(e^{-\eta/\tau} - e^{-t/\tau}) + \eta - t]\right|^{1/2},$$

I_0 – модифицированная, а J_0 – обычная функция Бесселя, t_0 – – положительный корень уравнения $g \cdot (1 - e^{-t/\tau}) - (t/\tau) = 0$;

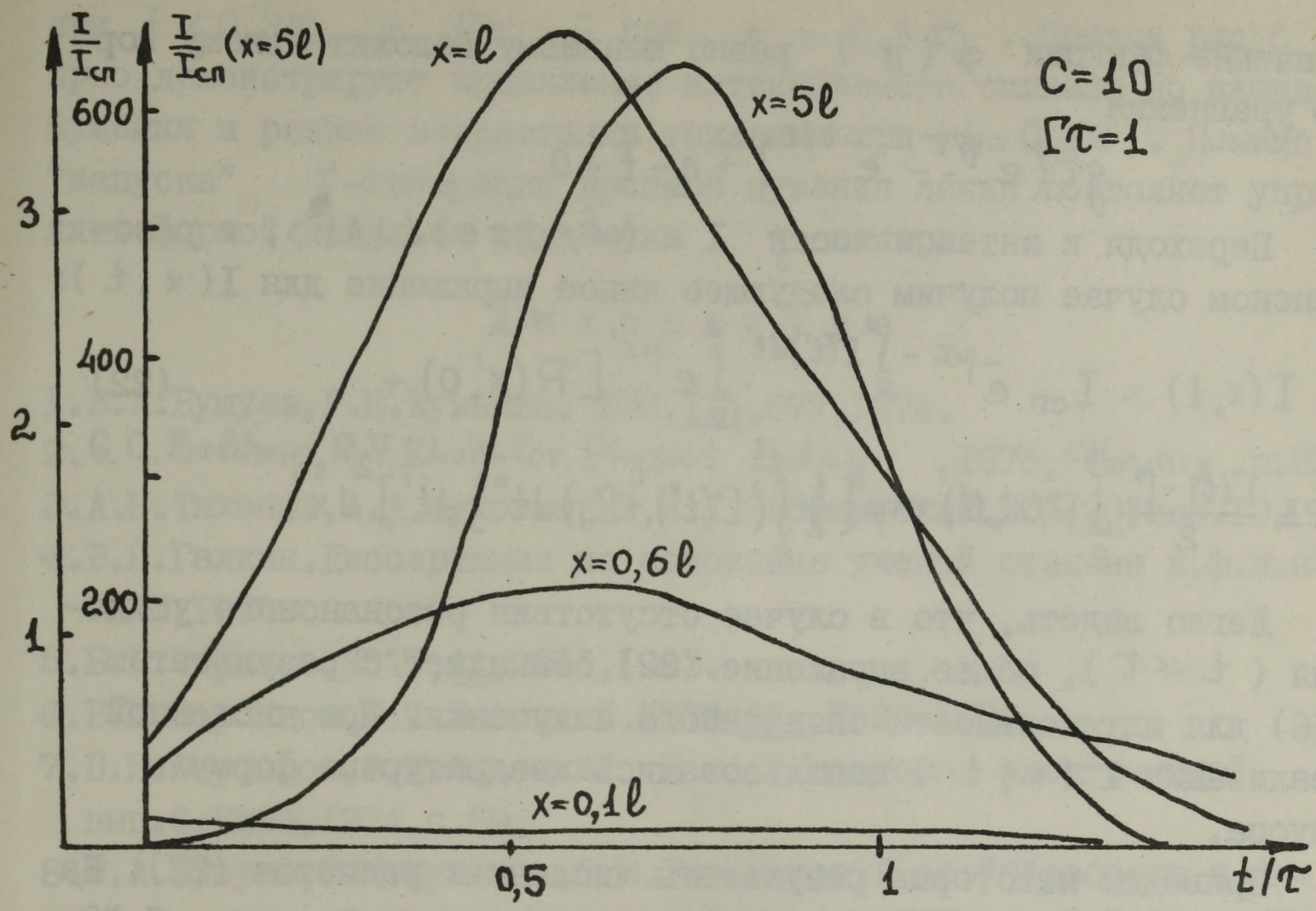


Рис. 5. Кинетика усиления спонтанного излучения ядер

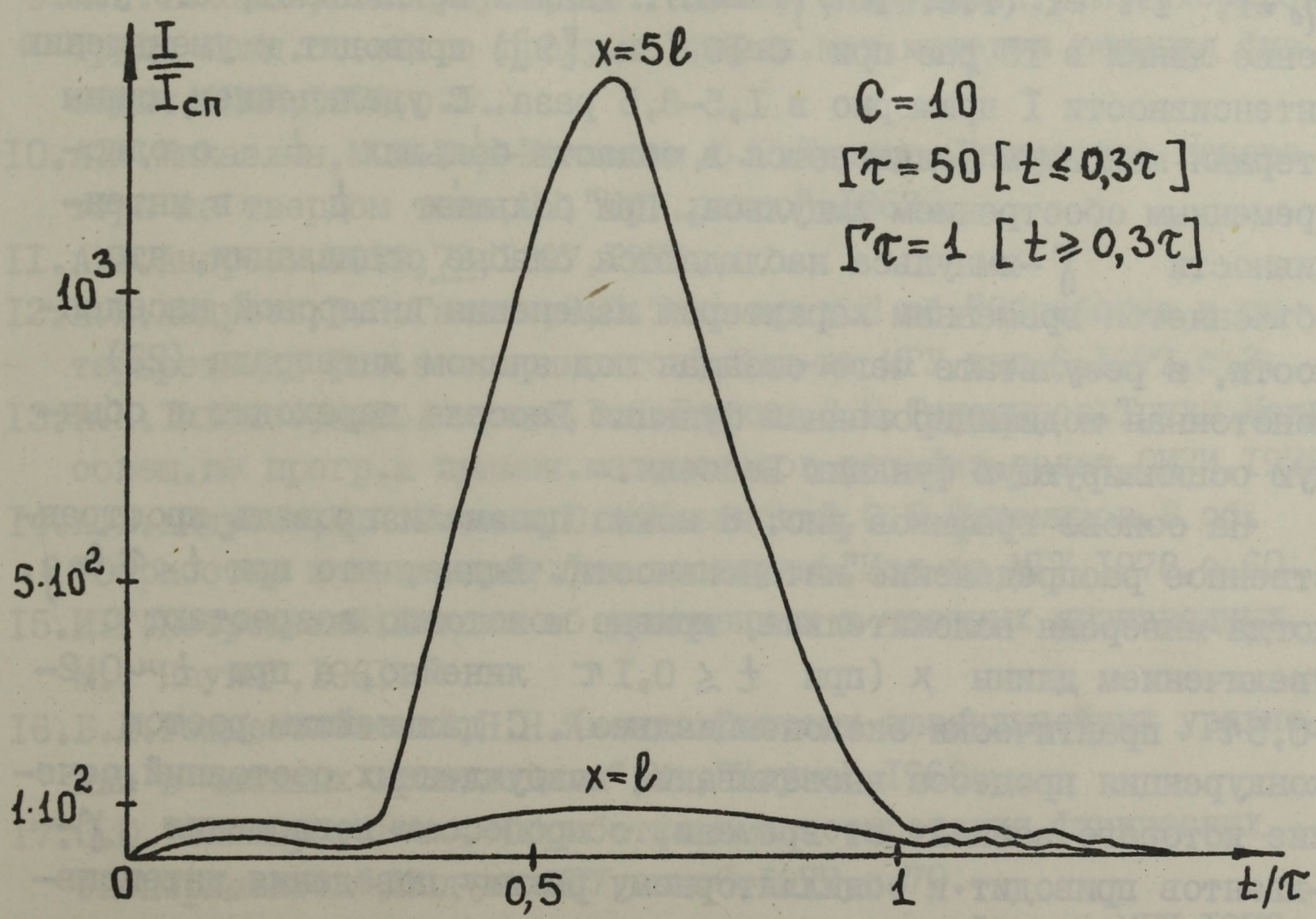


Рис. 6. Влияние процесса сужения линии $\Gamma(t)$ на характер кинетики усиленного спонтанного излучения ядер

значение функции $\Phi(\eta)$ равно меньшему положительному корню уравнения

$$g\tau\left(e^{-\eta/\tau} - e^{-t/\tau}\right) + \eta - t = 0.$$

Переходя к интенсивности $I = (\omega^2/2\pi c)\langle|A|^2\rangle$, в резонансном случае получим следующее явное выражение для $I(x,t)$:

$$I(x,t) = I_{сп}\, e^{-\mu x - \int_0^t \Gamma(t')dt'} \cdot \int_0^x e^{\mu x'} \Big[R(x',0) + \tag{22}$$

$$+ \frac{\Gamma(t)-\Gamma_0}{2} \int_0^t R(x',t') \exp\Big\{\frac{1}{2}\int_0^{t'} (\Gamma(t'')-\Gamma_0)dt''\Big\}dt'\Big]^2 dx'.$$

Легко видеть, что в случае отсутствия резонансного усиления ($t \ll \tau$), общее выражение (22) совпадает с результатом (I6) для интенсивности спонтанного излучения. При численной реализации $I(x,t)$ использовались квадратурные формулы Гаусса.

Приведем некоторые результаты численных расчетов (22). На рис. 5 показано временное развитие γ-волны в случае $C=10$, $\eta_0 = 1$, $\Gamma\tau = 1$ (т.е. $K/\mu = 40$). Анализ показывает, что уширение линии в 10 раз при $C=10$ (см. [9]) приводит к уменьшению интенсивности I примерно в 1,5–3,5 раза. С увеличением длины стержня максимум I сдвигается в область бо́льших t с одновременным обострением импульса. При бо́льших t в интенсивности γ-импульса наблюдаются слабые осцилляции, что объясняется временным характером изменения инверсной населенности, в результате чего стоящая под знаком интеграла (22) монотонная модифицированная функция Бесселя переходит в обычную осциллирующую функцию Бесселя.

На основе графиков рис. 5 можно проанализировать пространственное распределение интенсивности. Видно, что при $t < \tau \ln g$, когда инверсия положительна, кривые монотонно возрастают с увеличением длины x (при $t \lesssim 0{,}1\tau$ линейно, а при $t \sim 0{,}2$–$0{,}5\tau$ практически экспоненциально). С дальнейшим ростом конкуренция процесса высвечивания возбужденных состояний, сечение которого зависит от времени, с процессом поглощения γ-квантов приводит к осцилляторному режиму поведения интенсивности $I(x)$.

На рис. 6 представлены результаты расчета влияния процесса сужения неоднородной ширины Γ на временное поведение γ-импульса. Для $\Gamma(t)$ выбрана следующая модель сужения: $\Gamma\tau = 50$

при $t \leq 0{,}3\tau$ и $\Gamma\tau = 1$ при $t > 0{,}3\tau$. Кривые рис.6 ярко демонстрируют подавление интенсивности сигнала до начала сужения и резкое возрастание усиления при $t > 0{,}3\tau$. Помимо "запуска" γ-генерации процесс сужения линии позволяет управлять формой импульса излучения γ-лазера.

Литература

1. В.А.Бушуев, Р.Н.Кузьмин. УФН, 114, 677, 1974.
2. G.C.Baldwin, R.V.Khokhlov. Physics today, 1975, February, p.32.
3. А.Н.Тихонов, А.А.Жуховицкий, Я.Л.Забежинский. ЖФХ, 20, вып.10, 1946
4. В.Я.Галкин. Диссертация на соискание ученой степени к.ф.м.н. ОИЯИ, 1972.
5. Б.В.Чириков. ЖЭТФ, 44, 2016; 1963.
6. В.И.Воронцов, В.И.Высоцкий. ЖЭТФ, 66, 1528, 1974.
7. В.И.Воронцов, В.И.Высоцкий. В сб."Квантовая электроника", вып.8, Киев, 1974, с.69.
8. В.А.Бушуев, Р.Н.Кузьмин, О.Ю.Тихомиров. В сб."Обработка и интерпретация физ.экспериментов", Изд-во МГУ, вып.5, 1976, с.91.
9. А.Н.Тихонов, В.А.Бушуев, В.Я.Галкин, Р.Н.Кузьмин, О.Ю.Тихомиров. Труды Межд.совещ.по прогр.и примен.мат.методов решения физ. задач, ОИЯИ, 1978, с.5.
10. А.Л.Микаэлян, М.Л.Тер-Микаэлян, Ю.Г.Турков. Оптические генераторы на твердом теле, М., "Сов.радио", 1967.
11. А.В.Андреев. ЖЭТФ, 72, 1397, 1977.
12. А.В.Андреев, В.Я.Галкин, О,Ю.Тихомиров. В сб."Обработка и интерпретация физ.экспериментов", Изд-во МГУ, вып.6, 1977, с.3.
13. А.Н.Тихонов, А.В.Андреев, В.Я.Галкин, О.Ю.Тихомиров. Труды Межд. совещ.по прогр.и примен.мат.методов реш.физ.задач, ОИЯИ, 1978.
14. А.В.Андреев, В.Я.Галкин, Ю.А.Ильинский, О.Ю.Тихомиров. В сб. "Обработка и интерпрет.физ.эксперим.", Изд-во МГУ, 1978, с.60.
15. И.Г.Петровский. Лекции об уравнениях в частных производных, М., "Наука", 1961.
16. Б.Л.Рождественский, Н.Н.Яненко. Системы квазилинейных уравнений в частных производных", М., "Наука", 1968.
17. О.Ю.Тихомиров. В сб."Обработка и интерпретация физических экспериментов", Изд-во МГУ, вып.6, 1977, с.70.
18. О.Ю.Тихомиров. Диссер.на соиск.учен.степ.к.ф.м.н., МГУ, 1978.
19. А.Н.Тихонов, А.А.Самарский. Уравнения математической физики, М., "Наука", 1966.

НЕКОТОРЫЕ ВОПРОСЫ СУЩЕСТВОВАНИЯ И КАЧЕСТВЕННОГО ПОВЕДЕНИЯ ЧАСТИЦЕПОДОБНЫХ РЕШЕНИЙ

И.В.Амирханов, Е.П.Жидков

Объединенный институт ядерных исследований

Дубна

АННОТАЦИЯ

Рассмотрена краевая задача для нелинейного уравнения

$$\frac{d^2y}{dx^2} - Q_\ell(x)y = F(x,y), \qquad (1)$$

при граничных условиях

$$y(o) = y(\infty) = 0. \qquad (2)$$

В работе дан краткий обзор литературы, посвященной вопросам существования решений краевой задачи (1)-(2).

ABSTRACT

A boundary value problem for nonlinear equation

$$\frac{d^2y}{dx^2} - Q_\ell(x)y = F(x,y) \qquad (1)$$

for boundary conditions

$$y(0) = y(\infty) = 0 . \qquad (2)$$

is considered.

The review of references devoted the existence solution to the boundary value problem (1)-(2) is given.

Введение

Различные физические модели, используемые в таких разделах физики, как гидродинамика, плазма, физика твердого тела, нелинейная оптика, а также нелинейная теория поля, приводят к рассмотрению нелинейных уравнений /I-IO/.

В последнее время большое внимание со стороны физиков и математиков уделяется нелинейным уравнениям, имеющим так называемые солитонные решения. Это в первую очередь связано с тем, что в работе /II/ были обнаружены глубокие связи между уравнением Кортевега-де Фриза (КдФ)

$$u_t - 6\,u\,u_x + u_{xxx} = 0,$$

и спектральными свойствами семейства операторов Штурма-Лиувилля

$$-\frac{d^2}{dx^2} + u(x,t) \quad (-\infty < x < \infty),$$

порождаемых решениями $u(x,t)$ уравнения КдФ. Эти связи позволили найти частные (солитонные) решения уравнения КдФ с помощью обратной задачи теории рассеяния. Работа П.Лакса /I2/ оказала большое влияние на последующие исследования в этом направлении. Он ввел понятие операторной L-A -пары и показал, что уравнение КдФ эквивалентно операторному уравнению

$$\frac{dL}{dt} = [A, L],$$

характерной чертой которого является существование бесконечного числа первых интегралов. В.Е.Захаров и Л.Д.Фаддеев /I3/ установили, что уравнение КдФ является вполне интегрируемой гамильтоновой системой. Метод обратной задачи получил дальнейшее развитие в работах В.Е.Захарова и А.Б.Шабата /I4/. Следующий шаг в этом направлении сделан в работах С.П.Новикова /I5/, В.А.Марченко /I6/ и П.Лакса /I7/, где различными методами изучена периодическая задача для уравнения КдФ.

Практически все результаты, полученные для уравнения КдФ, были обнаружены /I-3/ впоследствии для большого числа интересных нелинейных уравнений. Все-таки следует отметить, что точные решения удается получать лишь в исключительных случаях, когда физическая модель является очень упрощенной.

Различные физические модели, используемые в теории элементарных частиц /4-9/, приводят к исследованию нелинейных уравнений в частных производных, поиск точного решения которых - практически безнадежная задача. Для нелинейной полевой теории элементарных частиц представляет большой интерес исследование существования и качественного поведения частицеподобных решений таких уравнений, что также являются сложной математической задачей.

Однако в некоторых случаях, используя определенные упрощения как физического, так и математического характера, удается перейти от уравнений в частных производных к обыкновенным нелинейным дифференциальным уравнениям. Тогда вопрос о существовании частицеподобных решений сводится к разрешимости краевых задач для этих уравнений. Так, например, в работе /I8/ рассматривается следующее уравнение

$$\angle y + F_1(y) = \lambda F_2(y), \qquad (I)$$

где $$L = -\sum a_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + a_0 , \quad i,j = 1,2,\dots,n,$$

(a_{ij}) - постоянная положительно определенная матрица,

$$a_0 > 0, \quad x = (x_1, \dots, x_n) \in R^n,$$

$F_1(y)$ и $F_2(y)$ - нелинейные функции, удовлетворяющие некоторым ограничениям. Используя метод симметризации, доказано, что если специально сформулированная вариационная задача имеет решение, то уравнение (I) имеет решение, котрое неотрицательно, радиально и экспоненциально затухает при $|x| \to \infty$. Тогда в дальнейшем вместо уравнения (I) можно исследовать радиальное уравнение

$$\ddot{u} - \frac{n-1}{r} \dot{u} - u - F_1(u) = -\lambda F_2(u), \qquad (2)$$

где $$r = |x| \neq 0, \ \frac{du}{dr} = \dot{u}, \ \frac{d^2u}{dr^2} = \ddot{u}.$$

В работе /I8/ для уравнения (2), используя результаты работы /I9/, доказано существование бесконечной последовательности различных радиальных решений.

Основная цель настоящей статьи состоит в выявлении условий разрешимости и свойств решений краевых задач, которые встречаются в полевой теории элементарных частиц. Эта цель определила отбор материала.

Некоторые полевые модели /4-9/ приводят к рассмотрению существования частицеподобного решения следующего нелинейного дифференциального уравнения

$$\ddot{\psi} + \frac{2}{x} \dot{\psi} - Q_\ell(x) \psi = F(\psi), \qquad (3)$$

где $$\ddot{\psi} = \frac{d^2\psi}{dx^2}, \ Q_\ell(x) = \frac{\ell(\ell+1)}{x^2} + \eta^2, \ \ell = 0,1,2,\dots,$$

$F(\psi)$ - некоторая нелинейная функция. Выбирая из физических соображений конкретный вид $F(\psi)$ - получаем различные модели /4-9/

После замены переменных $y(x)=x\psi(x)$ имеем

$$\ddot{y} - Q_\ell(x)y = x\,F\left(\frac{y}{x}\right), \tag{4}$$

$$y(0)=y(\infty)=0. \tag{5}$$

В данной работе под частицеподобным решением понимается любое нетривиальное решение $y(x)$ краевой задачи (4)-(5). Положительное частицеподобное решение - это частицеподобное решение, которое обращается в нуль только при $x=0$ и $x=\infty$, а в остальных точках интервала $(0,\infty)$ $y(x)>0$.

Исследование краевой задачи (4)-(5) проводились различными авторами в основном двумя методами. В основе одного из них лежит вариационный подход, а в основе другого - исследование качественного поведения решений дифференциальных уравнений.

Настоящая работа состоит из двух разделов.

В первом разделе краевая задача (4)-(5) исследуется вариационными методами, а во втором - методами качественной теории дифференциальных уравнений.

I. В этом разделе изложены вариационные методы изучения частицеподобных решений и даны их применения к исследованию разрешимости краевой задачи (4)-(5).

Ряд краевых задач был исследован в работах Нехари /20-22/. В работе /22/ приводится достаточное условие существования положительного частицеподобного решения уравнения

$$\ddot{y} - y = -\frac{y^k}{x^{k-1}}, \tag{I.1}$$

при граничных условиях

$$y(0)=y(\infty)=0. \tag{I.2}$$

Отметим, что уравнение (I.I) получается из уравнения (4) при $\ell = 0$ и $F\left(\frac{y}{x}\right) = -\left(\frac{y}{x}\right)^K$.

Задачу (I.I)-(I.2) Нехари сводит к решению изопериметрической задачи: найти минимум функционала

$$J(y) = \int_0^\infty \left[(\dot{y})^2 + y^2 \right] dx , \quad y(0) = y(\infty) = 0, \qquad (I.3)$$

при условии нормировки

$$K(y) = \int_0^\infty \frac{y^{K+1}}{x^{K-1}} dx = 1. \qquad (I.4)$$

Нехари показал, что из существования решения вариационной задачи (I.3)-(I.4) следует существование решения краевой задачи (I.I)-(I.2).

Этим методом доказано существование положительного частицеподобного решения задачи (I.I)-(I.2) при $1 < K < 5$. Кроме этого, показано неразрешимость этой задачи при $K = 5$.

В работах /23-24/ проведено дальнейшее исследование разрешимости краевой задачи (I.I)-(I.2) и доказана следующая

Теорема. Для любого целого положительного n $(n = 0, 1, 2, \ldots,)$ и любого $K = \frac{2P+1}{2q+1}$ (P и q - натуральные числа), $1 < K < 4$ существует решение $y(x)$ задачи (I.I)-(I.2), имеющее в точности n нулей на интервале $0 < x < \infty$. Была также доказана неразрешимость этой задачи для любого K : $K \geqslant 5$.

Работа /25/ посвящена вопросу о единственности положительных частицеподобных решений задачи (I.I)-(I.2).

В работе /26/ проведено исследование следующей краевой задачи

$$\ddot{y} - y = -y F_1(y^2, x), \qquad (I.5)$$

$$y(0) = y(\infty) = 0. \qquad (I.6)$$

Очевидно, что уравнение (I.5) получается из (4) при $\ell = 0$ и $xF = -y F_1(y^2, x)$.

Предполагается, что $F_1(u,x)$ удовлетворяет следующим условиям

а) $F_1(u,x)$ - непрерывна при $0<x<\infty$, $0\le u<\infty$;

б) $F_1(u,x)>0$, при $u>0$, $x>0$;

в) существует δ такое, что для любого фиксированного $x>0$ и $0\le u_1<u_2<\infty$,

$$u_2^{-\delta}F_1(u_2,x)>u_1^{-\delta}F_1(u_1,x).$$

Доказана следующая

Теорема. Если $F_1(u,x)$ удовлетворяет условиям (а)-(в) и кроме этого пусть $\lim\limits_{x\to\infty}F_1(c^2,x)=0$ для всех конечных $c>0$, и $\int\limits_0^a x^{\left(\frac{1}{2}-\varepsilon\right)}F_1(c^2x,x)dx<+\infty$ для всех конечных $c>0$, $0<a<\infty$ и некоторого $\varepsilon>0$, тогда уравнение (I.5) имеет бесконечное множество решений $\{y_n\}$, $n=1,2,\ldots$, непрерывных вместе со своими производными первых порядков в $[0,\infty)$ и функция $y_n(x)$ имеет точно $n-1$ нулей на $(0,\infty)$. Кроме того, $y_n(0)=0$, $\lim\frac{y_n(x)}{x}$ существует при $x\to 0$ и $y_n(x)\to 0$, когда $x\to\infty$ для каждого n.

Если нелинейную функцию $F\left(\frac{y}{x}\right)$ выбрать в виде

$$F\left(\frac{y}{x}\right)=-A\left(\frac{y}{x}\right)^k,$$

то из уравнения (4) получим

$$\ddot{y}-Q_\ell(x)y=-A\frac{y^k}{x^{k-1}}, \qquad (I.7)$$

$$y(0)=y(\infty)=0. \qquad (I.8)$$

Уравнение (I.7) при $\ell=0$ ранее исследовалост многими авторами /5,23-25,3I,33/.

В работе /27/ проведено исследование этого уравнения при любых натуральных значениях параметра ℓ. С помощью вариационного подхода, который является дальнейшим развитием метода Нехари /22/, доказана следующая основная

Теорема. При любых значениях параметров

$$A>0,\ \eta^2>0,\ \ell=0,1,2,\dots,$$

условие $1<K<5$ является достаточным для существования положительного частицеподобного решения краевой задачи (I.7)-(I.8).

Дальнейшее исследование уравнения (I.7) проведено в работах /28-29/, где доказаны теоремы об отсутствии положительного частицеподобного решения краевой задачи (I.7)-(I.8) при $0<K\leq 1$ и при $K\geqslant 5$.

В заключение этого раздела отметим, что сведения к вариационным задачам применялись для доказательства разрешимости краевых задач и в других работах /I8,I9,30/.

2. В этом разделе обсудим некоторые работы, где существование решений краевых задач исследуется методами качественной теории дифференциальных уравнений, а в некоторых случаях результаты получены численно на ЭВМ.

В работе /5/ уравнение (I.I) решалось численно на ЭВМ. Результаты счета указывали на то, что краевая задача (I.I)-(I.2) имеет частицеподобные решения с любым числом узлов. Краевая задача (I.I)-(I.2) рассматривалась и в работах /3I-32/. В них есть некоторые соображения относительно существования и свойств решения краевой задачи и приведены результаты машинного счета.

Затем в двух работах /22/ и /33/, вышедших одновременно, было строго доказано существование положительного частицеподобного решения краевой задачи (I.I)-(I.2). Работу /22/ обсуждали в первом разделе, поэтому здесь приведем некоторые основные результаты, полученные в работе /33/, а именно:

а) доказана теорема об отсутствии положительного частицеподобного решения при $0<K\leq 1$.

б) для уравнения (I.I) решение задачи Коши существует и единственно при $x>0$ и $K>1$.

в) доказана теорема существования положительного частицеподобного решения при $1<K\leq 3$.

Рассмотрение другой физической модели привело авторов /6,8/ к исследованию следующей краевой задачи

$$\ddot{\bar{y}}+\frac{2}{x}\dot{\bar{y}}-\eta^2\bar{y}=B_1\bar{y}^3+B_2\bar{y}^5, \qquad (2.1)$$

$$\dot{\bar{y}}(0)=\bar{y}(\infty)=0. \qquad (2.2)$$

В работе /6/ $B_1=-4, B_2=3$ и параметр η может принимать различные значения $\eta^2>0$, а в работе /8/ $\eta=1, B_1=-1$ и B_2 может принимать различные значения. Заметим, что заменой функции и переменного один случай можно свести к другому.

В работах /34/ приведены некоторые необходимые условия существования решений задачи (2.I)-(2.2).

Проводя качественное исследование уравнения (2.I) в фазовой плоскости и решая это уравнение численно на машине, авторы /8/ пришли к выводу, что при любом $-\infty<B_2<\frac{3}{16}$ задача (2.I)-(2.2) имеет частицеподобное решение. На основании же результатов счета на ЭВМ и некоторых нестрогих рассуждений авторы /6/ утверждали то же самое, при $0<\eta<1$. Изложим кратко суть этих рассуждений.

Для этого рассмотрим уравнение (2.I) с начальными условиями $\bar{y}(0)=y_0$, $\dot{\bar{y}}(0)=0$, т.е. задачу Коши. Исследуем решение задачи Коши на фазовой плоскости $\bar{y}, \dot{\bar{y}}$. Уравнение (2.I) без члена $\frac{2}{x}\dot{\bar{y}}$ имеет вид:

$$\ddot{y}-\eta^2 y=B_1 y^3+B_2 y^5 \qquad (2.3)$$

Для него нетрудно получить картину фазовых траекторий.

Первый интеграл уравнения (2.3) (закон сохранения энергии) имеет вид

$$T(\dot{y}) + V(y) = E(y_0), \tag{2.4}$$

где

$T(\dot{y}) = \frac{1}{2}(\dot{y})^2$ - кинетическая энергия,

$V(y) = -\frac{1}{2}\eta^2 y^2 - \frac{1}{4}B_1 y^4 - \frac{1}{6}B_2 y^6$ - потенциальная энергия,

$E(y_0) = V(y_0)$ - постоянное интегрирование

Точки равновесия находим из уравнения

$$\frac{\partial V}{\partial y} = 0. \tag{2.6}$$

Например, если $B_1 = -4$, $B_2 = 3$ и $0 < \eta^2 < 1$, то качественная картина фазовых траекторий и график потенциальной энергии $V(y)$ приведены на рис. I.

Вернемся к уравнению (2.I). Здесь присутствует член $\frac{2\dot{y}}{x}$, вследствие чего полная энергия не сохраняется, т.е.

$$\frac{dE(\bar{y})}{dx} = -\frac{2(\dot{y})^2}{x} \tag{2.7}$$

в отличие от (2.4), где $E = const$. Следовательно, везде $\frac{\partial E}{\partial x} < 0$ и энергия монотонно убывает. Это исключает возможность существования замкнутых предельных циклов, и положением равновесия может быть только одна из точек $0, \pm y_3$ (неустойчивое) и $\pm y_1$ (устойчивое). Поэтому на фазовой плоскости любая траектория, соответствующая решениям уравнения (2.I), будет пересекать линии $E = const$ и заканчиваться в одной из точек $0, \pm y_1$. Если траектория пересечет линию $E = 0$ в правой полуплоскости, то она закончится в точке y_1, а если пересечение произойдет в левой полуплоскости, то траектория закончится в точке $-y_1$. Решениям, удовлетворяющим граничным условиям (2.2), соответствуют траектории, кон-

чающиеся в точке О. Эти траектории заключены между траекториями, заканчивающимися в y_1 и $-y_1$.

Благодаря начальным условиям $\dot{y}(0)=0$, $y(0)=y_0$, мы интересуемся лишь теми траекториями, которые начинаются в случае фазовой плоскости $y, \dot{y}$ на оси y. Поэтому траектории, заканчивающиеся в точке О, следует искать путем плавного увеличения начальной амплитуды y_0 на оси y. Когда $0<y_0<y_2$, все траектории кончаются в $+y_1$. В окрестности $+y_1$ для y_0 несколько больших картина не изменится. Траектория, начинающаяся в точке y_{11}, показана на рис.I. Когда y_0 возрастает до y_{22}, потенциальная энергия V увеличивается настолько, что траектория пересекает кривую $E=0$ в левой полуплоскости и кончается в точке $-y_1$. Траектория, кончающаяся в точке О, находится между двумя траекториями, которые начинаются в точках y_{11} и y_{22}. Заметим, что доказательство существования таких начальных значений $\dot{y}(0)=0$ и $y(0)=y_{22}$, при которых траектория пересекает кривую $E=0$ в левой полуплоскости, является сложной задачей для уравнения (2.I). Без строгого доказательства этого факта, установление существования частицеподобных решений задачи (2.I)-(2.2), как это делается в работах /6,8/, является не строгим результатом. Нетрудно привести противоречащий пример. Для уравнения

$$\ddot{y}+\frac{2}{x}\dot{y}-y=-y^5 \qquad (2.8)$$

качественная картина фазовых траекторий и график потенциальной энергии $V(y)$ приведены на рис.2.

Из рис.2. видно, что, повторяя все качественные рассуждения, приведенные вышеуказанными авторами /6,8/, можно было бы заключить, что уравнение (2.8) при граничных условиях $\dot{y}(0)=0$ и $y(\infty)=0$

имеет решение. Но как строго доказано в работах /22,33/, эта задача не имеет решение.

Причина противоречия в первую очередь связана с тем, что с увеличением y_0 , хотя возрастает потенциал $V(y)$, но все траектории заканчиваются в точке +I. Кроме этого, в качественных рассуждениях использовали непрерывную зависимость задачи Коши от начальных амплитуд y_0 в интервале $0<x<\infty$, что не всегда верно для произвольного нелинейного уравнения.

Строгое исследование краевой задачи (2.I)-(2.2) методом качественной теории обыкновенных дифференциальных уравнений проведено в работе /36/. Строго доказано существование положительных частицеподобных решений для всех малых значений параметра γ.

В работе /35/ изучается краевая задача для обыкновенного дифференциального уравнения второго порядка.

$$\ddot{y}+\frac{2}{x}\dot{y}=f(y), \qquad 0\leq x<+\infty, \tag{2.9}$$

$$\dot{y}(0)=0, \quad y(+\infty)=0. \tag{2.10}$$

Доказано существование частицеподобных решений для довольно широкого класса функций $f(y)$. В частности, доказаны теоремы существования частицеподобных решений с n узлами.

Литература

I. A.Scott, F.Chu, D.McLaughlin.Proc. of IEE.E, 61, 1443, 1, 1973.

2. Дж.Уизем. Линейные и нелинейные волны. М., "Мир", I977.

3. V.G.Makhankov. Phys. Reports, 35C, 1, 1978.

4. R.J.Finkelstein, R.Lelevier and M.Ruderman.
Phys. Rev. , 83, 326, 1951.

5. В.Б.Гласко, Ф.Лерюст, Я.П.Терлецкий, С.Ф.Шушурин. ЖЭТФ, 35, 452, I958.

6. R.Friedberg, T.D.Lee, A.Sirlin.
Nucl. Phys. B115 , 1, 1976; B115, 32, 1976.

7. Л.Г.Заставенко. ПММ, 29, 430, I965.

8. D.L.T. Anderson. J.Math. Phys. 12, 945, 1971.

9. G. Rosen. J.Math. Phys. 7, 2066, 1966 .

I0. В.В.Бабиков. Препринт ОИЯИ Р4-4248, Дубна, I969.

II. C.S. Gardner, I.M. Green, M.D. Kruskal, R.M. Miura.
Phys. Rev. Letters, 19, 1095, 1967 .

I2. P.D.Lax. Comm. Pure. Appl. Math. 21, No 2, 467, 1968.

I3. В.Е.Захаров, Л.Д.Фаддеев. Функциональный анализ, 5, вып.4, I8, I97I.

I4. В.Е.Захаров, А.Б.Шабат. ЖЭТФ, 6I, II8, I97I; 64, I627, I973; Функциональный анализ 8, вып. 3, 43, I974.

I5. С.П.Новиков. Функциональный анализ, 8, вып. 3, 54, I974.

I6. В.А.Марченко. Математический сборник, 95 (I37), вып.3, 33I, I974.

I7. P.D.Lax. Lectures Appl. Math. 15, 85, 1974.

I8. W.A. Strauss. Commun. Math. Phys. 55, 149, 1977.

I9. P.H.Rabinowitz. Indiana U.Math. J. 23, 729, 1974.

20. Z.Nehari. Trans. Amer. Math. Soc., 95, 101, 1960 .

2I. Z.Nehari. Acta Math. 105, 141, 1961.

22. Z.Nehari. Proc. Royal Irish. Acad., A62, 117, 1963.

23. В.П.Шириков. Препринт ОИЯИ Р-I682, Дубна, I864; ДАН СССР, I63, 834, I965.

24. Е.П.Жидков, В.П.Шириков, И.В.Пузынин. Препринт ОИЯИ 2005, Дубна, I965.

25. В.П.Шириков. Препринт ОИЯИ 2006, Дубна, I965.

26. G.H.Ryder. Pacific J. Math. 22, 477, 1967.

27. И.В.Амирханов, Е.П.Жидков, Г.И.Макаренко. Сообщение ОИЯИ Р5-II705, Дубна, I978.

28. И.В.Амирханов, Г.И.Макаренко. Сообщение ОИЯИ Р5-II865, Дубна, I978.

29. И.В.Амирханов, Е.П.Жидков, Г.И.Макаренко. Сообщение ОИЯИ Р5-II866, Дубна, I978.

30. M.S.Berger. J.Funct. Anal. 9, 249, 1972.

3I. I.L.Singe. Proc. Royal. Irish. Acad. A62, 17, 1961.

32. N.Rosen, H.B.Rosenstock. Phys. Rev., 82, 257, 1952.

33. Е.П.Жидков, В.П.Шириков. Препринт ОИЯИ Р-I3I9, Дубна, I963; ЖВМиМФ, 4, 804, I964.

34. V.G.Makhankov. Phys. Lett. 61A, 431, 1977.

Ю.В.Катышев, В.Г.Маханьков. Сообщение ОИЯИ Р2-I0547, Дубна, I977.

35. Е.П.Жидков, П.Е.Жидков. Сообщение ОИЯИ Р5-11599, Дубна, 1978.

36. Е.П.Жидков, П.Е.Жидков. Сообщение ОИЯИ Р5-11600, Дубна, 1978.

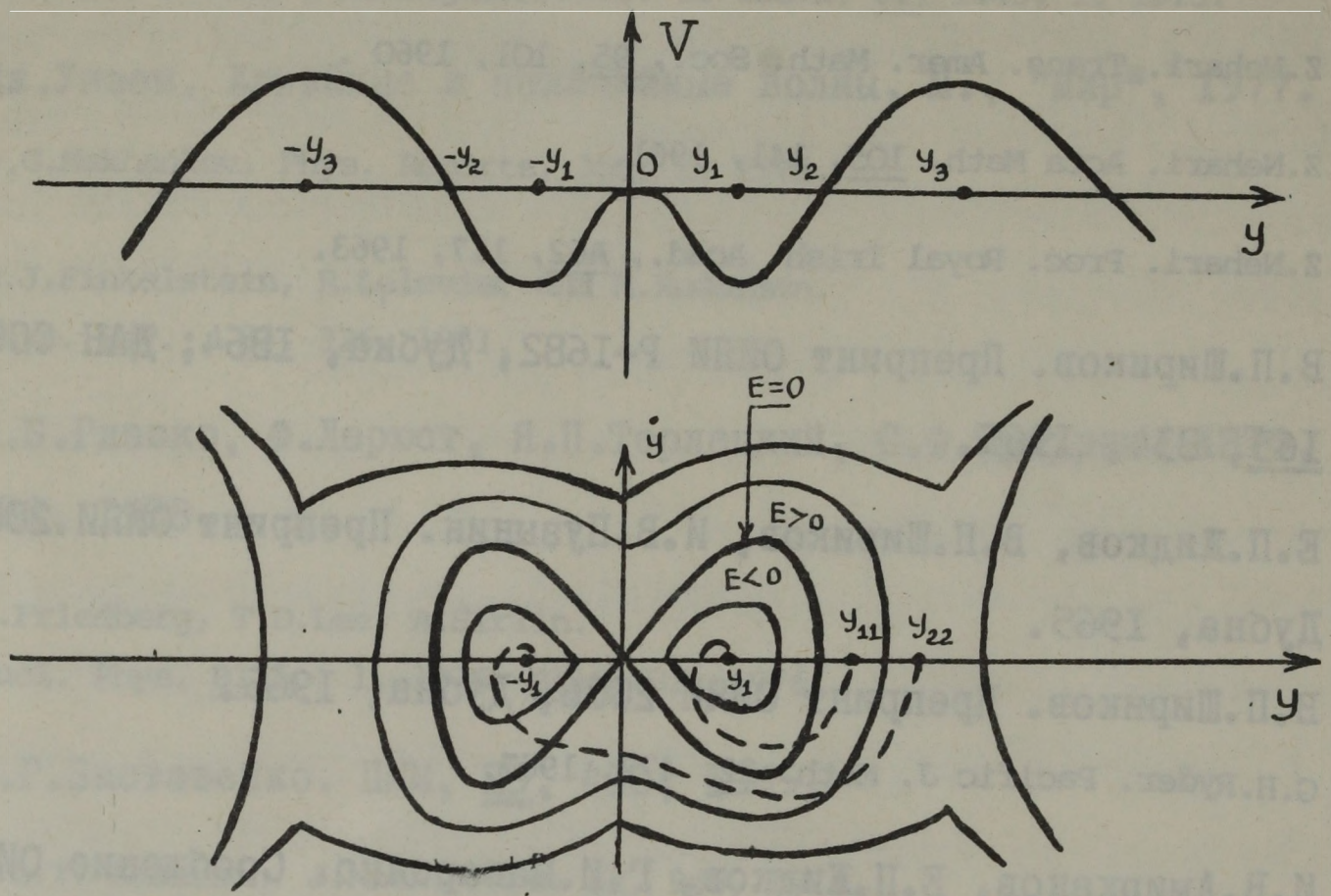


Рис. I.

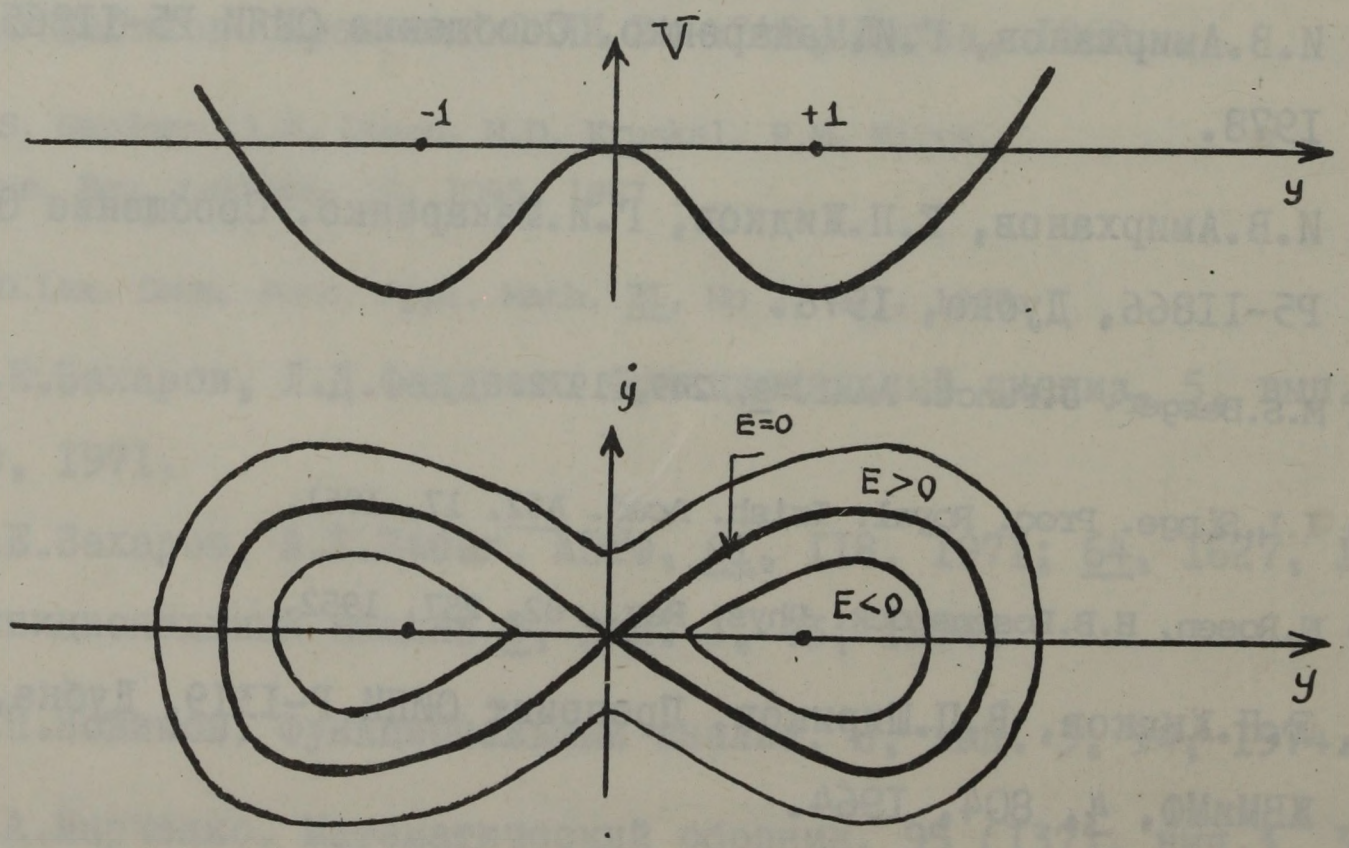


Рис. 2.

О РЕШЕНИИ КВАЗИЛИНЕЙНОЙ ПАРАБОЛИЧЕСКОЙ СИСТЕМЫ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ В ЦИЛИНДРЕ

Р.Х.Фарзан, Д.Молнарка
Будапештский Университет им. Лоранда Этвеша,
Будапешт

АННОТАЦИЯ

Для истемы квазилинейных дифференциальных уравнений параболического типа в цилиндре с осевой симметрией, моделирующей процессы в химических реакторах, строятся разностные схемы второго порядка точности. Доказываются теоремы о достаточных условиях сходимости и устойчивости схемы в пормах, эквивалентных равномерной, а также о сходимости итерационного процесса решения нелинейной системы.

ABSTRACT

Discrete approximations with second order accuracy for quasilinear parabolic differential equations are considered. For these model equations of chemical reactor in cylindrical geometry with axial symmetry theorems are proved on sufficient conditions of convergence and of stability.

В настоящей работе результаты, полученные в [I] для одного уравнения, обобщаются на случай системы дифференциальных уравнений. В работе [2] авторы провели такое обобщение для одномерного параболического уравнения.

I. При исследовании процессов, протекающих в химических реакторах, встречаются с необходимостью численного решения параболической системы дифференциальных уравнений со слабой нелинейностью вида:

$$/I/ \qquad \frac{\partial u_k}{\partial t} = \operatorname{div}(D_k \operatorname{grad} u_k) - \operatorname{div}(\underline{v}\, u_k) + f_k(\underline{r}, t, u_1, \ldots u_K), \qquad k = 1, \ldots K ;$$

где u_k – плотность компонент, присутствующих в потоке, $D_k(\underline{r},t)$ – коэффициент диффузии, $\underline{v}(\underline{r},t)$ – вектор скорости потока, f_k – функция источника.

В данной работе исследуется численное решение этой системы уравнений в случае осесимметрического потока в цилиндре, когда входящие в уравнение /I/ функции /а следовательно и само решение/ в пространственных цилиндрических координатах $\{r, \varphi, z\}$ зависят только от r и z. В этом случае система уравнений /I/ перепишется в виде:

$$\frac{\partial u_k}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left(rD_k\frac{\partial u_k}{\partial r}\right) - \frac{1}{r}\frac{\partial}{\partial r}(rwu_k) + \frac{\partial}{\partial z}\left(D_k\frac{\partial u_k}{\partial z}\right) -$$

/2/
$$- \frac{\partial}{\partial z}(vu_k) + f_k ,$$

$$\{r,z,t\} \in Q = \Omega \times (0,T], \quad \Omega = \{r,z \mid 0 < r < R,\ 0 < z < Z\};$$

где v и w – осевая и радиальная составляющие вектора скорости $\underline{v}$. Относительно функций, входящих в /2/ предполагаем:

/3/
$$0 < D_k(r,z,t) \in C^{110},$$
$$v(r,z,t) \in C^{010},$$
$$w(r,z,t) \in C^{100},$$
$$f_k \in C .$$

Для f_k кроме того предполагаем выполненными условия Липшица по всем u_m :

/4/
$$\left|f_k(r,z,t,u_1^*,\ldots,u_K^*) - f_k(r,z,t,u_1^{**},\ldots,u_K^{**})\right| \le \sum_{m=1}^{K} L_k^m \left|u_m^* - u_m^{**}\right| .$$

Для системы уравнений /2/ задаются начальные условия и краевые условия смешанного типа:

$$u_k(r,z,0) = U_k(r,z), \quad (r,z) \in \bar{\Omega} ;$$

/5/
$$u_k(r,0,t) = \varphi_k^{(1)}(r,t), \quad u_k(r,Z,t) = \varphi_k^{(2)}(r,t),$$
$$r \in [0,R], \quad t \in (0,T] ;$$
$$\frac{\partial u_k}{\partial r} - \sigma_k u_k = \psi_k(z,t), \quad \sigma_k(z) \geqslant 0, \quad r = R, \quad z \in (0,Z), \quad t \in (0,T].$$

Через $\bar{\Omega}$ будем обозначать замыкание области Ω . В области $\bar{\Omega}$ введем сетку $\bar{\Omega}_h$:

$$\bar{\Omega}_h = \Big\{ r_j, z_i \;/\; r_j = (j+\tfrac{1}{2})h_r,\ j = 0, 1, \ldots, J\ ,$$
$$h_r = \frac{R}{J+\frac{1}{2}}\ ;\quad z_i = ih_z,\ i = 0, \ldots, I\ ,\ h_z = \frac{Z}{I} \Big\}.$$

Сетку на всей области $\bar{Q}$ определим как

$$\bar{Q}_h = \Big\{ r_j, z_i, t^n \;/\; (r_j, z_i) \in \bar{\Omega}_h\ ;\ t^n = n\tau\ ,\ n = 0, \ldots, N\ ,\ \tau = \frac{T}{N} \Big\}$$

/В дальнейшем в некоторых случаях, не вызывающих недоразумений, для краткости будем писать: $(j,i) \in \bar{\Omega}_h$ /

На сетке определим сеточные функции $B^n_{k,ji} = B_k(r_j, z_i, t^n)$, причем при написании разностной схемы для коэффициентов будут использоваться как целые, так и полуцелые индексы.

Введем определение вектора не сетке. Пусть

$$y^n_k = \big\{ y^n_{k,ji} \big\},\quad (j,i) \in \bar{\Omega}_h\ ,$$
/6/
$$f^n_k = \big\{ f_k(r_j, z_i, t^n, y^n_{1,ji}, \ldots, y^n_{K,ji}) \big\},\quad (j,i) \in \bar{\Omega}_h\ .$$

Таким образом, $y^n_k \in Y_k \subset \mathbb{R}^{(J+1)\times(I+1)}$, $f^n_k \in F_k \subset \mathbb{R}^{(J+1)\times(I+1)}$.

Исследования в настоящей работе проводятся в равномерной метрике с "нормой на слое". Для y^n_k используется норма

$$/7/ \qquad \| y^n_k \|_n = \max_{(j,i) \in \bar{\Omega}_h} | y^n_{k,ji} |\ .$$

Используя обозначения [3], запишем разностные операторы, аппроксимирующие линейные дифференциальные эллиптические операторы в правой части /2/ :

$$A^n_{k,z} y^n_k = \Big(D_k(r, z - \tfrac{h_z}{2}, t^n)(y^n_k)_{\bar{z}} \Big)_z - \frac{1}{2}\Big(v(r, z - \tfrac{h_z}{2}, t^n)\cdot(y^n_k)_{\bar{z}} +$$
/8/
$$+ v(r, z + \tfrac{h_z}{2}, t^n)\cdot(y^n_k)_z \Big)\ ;$$

$$A^n_{k,r}y^n_k = \frac{1}{r}\left((r-\frac{h_r}{2})D_k(r-\frac{h_r}{2},z,t^n)(y^n_k)_{\overline{r}}\right)_r -$$

$$-\frac{1}{2r}\left((r-\frac{h_r}{2})w(r-\frac{h_r}{2},z,t^n)(y^n_k)_{\overline{r}} + (r+\frac{h_r}{2})w(r+\frac{h_r}{2},z,t^n)(y^n_k)_r\right).$$

Такая аппроксимация является некоторым обобщением результатов [4]. Отметим, что операторы $A^n_{k,z}$ и $A^n_{k,r}$ несамосопряженные.

Запишем теперь систему разностных уравнений, аппроксимирующую систему дифференциальных уравнений /2/ :

/9/ $$\left(E - A^n_{k,z} - A^n_{k,r}\right)y^n_k = y^{n-1}_k + \tau f^n_k, \quad k = 1,\dots,K,$$
$$n = 1,\dots,N.$$

Разностные уравнения /9/ аппроксимируют /после деления на τ / исходные уравнения /2/ с точностью

$$O\left(\tau + \frac{h_r^2}{r} + h_z^2\right).$$

Та же точность сохраняется, если переписать /9/ в виде

$$\Lambda^n y^n_k = y^{n-1}_k + \tau f^n_k,$$

/10/ где

$$\Lambda^n_k = \Lambda^n_{k,z}\Lambda^n_{k,r}, \Lambda^n_{k,z} = E - \tau A^n_{k,z}, \Lambda^n_{k,r} = E - \tau A^n_{k,r}.$$

Таким образом, в отличие от /9/, где использован пятиточечный шаблон в $\overline{\Omega}_h$, здесь используется девятиточечный.

Начальные условия для /10/ получаются из /5/ :

/11/ $y^o_k = u^o_k$, где $u^o_k = \{U_k(r_j,z_i)\}$, $(j,i)\in\overline{\Omega}_h$

Можно включить граничные условия в /10/, т.е. распространить разностные операторы /8/ и на граничные точки области [1]. Перепишем уравнение /10/ в виде

/I2/ $$\Lambda^n_k y^n_k = \Delta y^{n-1}_k + \tau \tilde{f}^n_k$$

где

$$(\Delta y^n_k)_{ji} = \begin{cases} y^n_{k,ji}, & \text{если } i \neq 0 \text{ и } j \neq 0, \\ 0, & \text{если } i = 0 \text{ или } j = 0; \end{cases}$$

и $\tilde{f}^n_k = f^n_k$ во внутренних точках области.

Предположим:

$$(A^n_{k,z} y^n_k)_{j0} = (A^n_{k,r} y^n_k)_{j0} = 0, \quad \tilde{f}^n_{k,j0} = \frac{1}{\tau} \varphi^{(1)}_k(r_j, t^n),$$

$$(A^n_{k,z} y^n_k)_{jI} = (A^n_{k,r} y^n_k)_{jI} = 0, \quad \tilde{f}^n_{k,jI} = \frac{1}{\tau} \varphi^{(2)}_k(r_j, t^n),$$

/I3/

$$(A^n_{k,r} y^n_k)_{oi} = \left(\frac{2}{h_r} D^n_{k,1/2i} - w^n_{1/2i}\right)(y^n_{k,oi})_r, \quad \tilde{f}^n_{k,oi} = f^n_{k,oi},$$

$$(A^n_{k,r} y^n_k)_{Ji} = -\left(\frac{2}{h_r} D^n_{k,J-1/2i} \frac{r_{J-1/2}}{r_J} + w^n_{Ji}\right)(y^n_{k,Ji})_{\bar{r}} -$$

$$- \frac{2}{h_r}\left(\frac{r_J w_{Ji} - r_{J-1/2} w_{J-1/2i}}{r_J} - \sigma(z_i) D^n_{k,Ji}\right) y^n_{k,Ji},$$

$$\tilde{f}^n_{k,Ji} = f^n_{k,Ji} + \frac{2}{h_r} D^n_{k,Ji} \psi_k(z_i, t^n).$$

При $j = 0$ и $j = J$ оператор $\Lambda^n_{k,z}$ сохраняет вид /8/.

Заметим, что с помощью /I3/ граничные условия /5/ аппроксимируются с той же точностью

$$o\left(\tau + \frac{h_r^2}{r} + h_z^2\right).$$

Таким образом, система уравнений /I2/ аппроксимирует систему уравнений /2/ и граничные условия /5/ с одинаковой точностью во всех точках $\bar{Q}_h$ /кроме $n = 0$, для которого выписано /II/ /.

В дальнейшем будем опускать тильду у функции $\tilde{f}_k$.

Введем вектор на сетке, охватывающий все компоненты:

$$y^n = \{ y_1^n, \ldots, y_K^n \}, \quad y^n \in Y \subset \mathbb{R}^{K \times (J+1) \times (I+1)}.$$

Аналогично можно ввести вектор f^n

Определим K - норму вектора y^n :

/I4/ $$\| y^n \|_K = \sum_{k=1}^{K} \| y_k^n \|_n .$$

С помощью вектора y^n перепишем систему уравнений /I2/ :

/I5/ $$\Lambda^n y^n = \Delta y^{n-1} + \tau f^n , \quad n = 1, \ldots N .$$

Очевидно, Λ^n имеет блочную структуру, где вдоль главной диагонали расположены квадратные матрицы Λ_k^n размерностью $(I+1)(J+1) \times (I+1)(J+1)$. Остальные элементы нули.

Система уравнений /I5/ решается последовательно для $n = 1,2,\ldots,N$ с использованием уже известных значений искомых функций на предыдущем временном слое t^{n-1} причем y^0 известно из /II/.

На каждом временном слое предлагается решать уравнения /I5/ методом итераций, линеаризируя уравнения

/I6/ $$\Lambda^n \overset{(s+1)}{y}{}^n = \Delta y^{n-1} + \tau f^n \left(\overset{(s)}{y}{}^n \right) , \quad s = 0,1,\ldots ,$$

задавая в качестве $\overset{(0)}{y}{}^n$ например значение y^{n-1} на предыдущем слое.

В силу отмеченной выше блочной структуры линейного оператора Λ^n система линеаризированных уравнений распадается на подсистемы для отдельных k . Далее, очевидно, для каждого k и n оператор расслаивается, т.е. можно последовательно решать системы:

/I7/ $$\Lambda^n_{k,z}\, {}^{(s)}_{x}{}^{n}_{k} = {}^{(s)}_{F}{}^{n}_{k} = \Delta y^{n-1}_k + \tau f^n_k({}^{(s)}_{y}{}^{n}),$$

/I8/ $$\Lambda^n_{k,r}\, {}^{(s+1)}_{y}{}^{n}_{k} = {}^{(s)}_{x}{}^{n}_{k},$$

причем очевидно, операторы обеих систем можно представить в виде трехдиагональной матрицы. Следовательно, для решения /I7/, /I8/ применим метод прогонки.

2. Целью настоящей работы является исследование условий, накладываемых на параметры сетки τ, h_r и h_z для того, чтобы:

а/, решение сеточной краевой задачи /I5/, /II/ сходилось к решению исходной дифференциальной задачи /2/, /5/ ;

б/, итерационный процесс рещения квазилинейной системы /I6/ был сходящимся ;

в/, метод прогонки для решения систем /I7/, /I8/ был устойчивым.

Рассмотрим эти проблемы в обратном порядке. Перепишем уравнение /I7/, опуская индексы n и s:

/I9/ $$(\Lambda_{k,z} x_k)_{ji} = -\frac{\tau}{h_z^2}\left(D_{k,ji-1/2} + \frac{h_z}{2} v_{ji-1/2}\right)x_{k,ji-1} +$$

$$+\left[1 + \frac{\tau}{h_z^2}\left(D_{k,ji-1/2} - \frac{h_z}{2} v_{ji-1/2} + D_{k,ji+1/2} + \frac{h_z}{2} v_{ji+1/2}\right)\right] x_{k,ji} -$$

$$-\frac{\tau}{h_z^2}\left(D_{k,ji+1/2} - \frac{h_z}{2} v_{ji+1/2}\right)x_{k,ji+1/2} = F_{k,ji}\,, \quad 0 < i < I\,,$$

$$x_{k,j0} = \varphi^{(1)}_k(r_j,t), \quad x_{k,jI} = \varphi^{(2)}_k(r_j,t).$$

Для устойчивости метода прогонки для системы уравнений

$$a_i x_{i-1} + b_i x_i + c_i x_{i+1} = F_i$$

достаточно выполнения условий [3]:

/20/ $|b_i| - |a_i| - |c_i| \geq 0, \; i = 1,\ldots,N-1$.

Предположим, что

$$D_{k,ji-1/2} + \frac{h_z}{2} v_{ji-1/2} > 0,$$

$$D_{k,ji+1/2} - \frac{h_z}{2} v_{ji+1/2} > 0.$$

Для этого достаточно h_z выбрать так, чтобы выполнялось условие

/2I/ $$h_z \leq \min_{(r,z,t) \in \overline{Q}} \frac{2D_k(r,z,t)}{|v(r,z,t)|} .$$

При этом коэффициент при $x_{k,ji}$ в /I9/ будет больше единицы.

Теперь для выполнения условия /20/ для системы уравнений /I7/ или /I9/ при всех n достаточно потребовать:

/22/ $$\tau < L_v^{-1}, \quad L_v = \max_{\overline{Q}} |v_z'|.$$

Очевидно, условия /2I/, /22/ как правило легко выполнимые.

Аналогичные достаточные условия для сходимости метода прогонки для уравнения /I8/ имеют вид:

/23/ $$h_r < \min_{\overline{Q}} \frac{2D_k}{|w|}, \; \tau < L_w^{-1}, \; L_w = \max_{\overline{Q}} \max_{\substack{|\theta| \leq 1 \\ 0 < r + \frac{\theta h_r}{2} < R}} \frac{|rw_r'|}{r + \frac{\theta h_r}{2}} .$$

В дальнейшем нам понадобится оценка для норм операторов $\left(\Lambda^n_{k,z}\right)^{-1}$ и $\left(\Lambda^n_{k,r}\right)^{-1}$.

Обратимся к уравнению /I9/. Пусть j_o , i_o таковы, что для фиксированного n :

$$|x_{k,j_o i_o}| = \max_{\bar{\Omega}_h} |x_{k,ji}|.$$

Если такая точка не одна, то выбираем точку с наименьшими индексами. Без потери общности можно предположить, что

/24/
$$0 < i_o < I,$$
$$x_{j_o i_o} > 0.$$

Тогда, если в уравнении /I9/ для $j = j_o$ и $i = i_o$ заменить $x_{k,j_o i_o-1}$ и $x_{k,j_o i_o+1}$ на $x_{k,j_o i_o}$, то в силу отрицательности их коэффициентов получим неравенство:

/25/
$$\left(1 + \tau \frac{v_{j_o i_o+1/2} - v_{j_o i_o-1/2}}{h_z}\right) x_{k,j_o i_o} \le F_{k,j_o i_o},$$

где вследствие /22/ выражение в скобках положительно. Так как

$$x_k = (\Lambda_{k,z})^{-1} F_k, \quad \|F_k\|_n \ge F_{k,j_o i_o},$$

то из /25/и /22/ следует

/26/
$$\|(\Lambda^n_{k,z})^{-1}\|_n \le \frac{1}{1 - \tau L_v},$$

где норма матрицы соответствует равномерной норме вектора /7/. Далее отметим, что ограничения /24/ снимаются тривиально.

Совершенно аналогично для оператора $(\Lambda^n_{k,r})^{-1}$ получим:

/27/
$$\|(\Lambda^n_{k,r})^{-1}\|_n \le \frac{1}{1 - \tau L_w}.$$

3. Рассмотрим достаточные условия, при которых сходится итерационный процесс /I6/. Итерационный процесс

$$z^{(s+1)} = g(z^{(s)}), \quad s = 0,1,...$$

сходится при любом $z^{(0)} \in [z_1, z_2]$, [5], если существует $q < 1$ такое, что при любых $z^*, z^{**} \in [z_1, z_2]$ выполняется

/28/ $$\rho(g(z^*), g(z^{**})) \leq q\,\rho(z^*, z^{**}).$$

Определим расстояние ρ посредством нормы /I4/. Пусть $z \in \mathbb{R}^{K\times(J+1)\times(I+1)}$ и пусть

$$\rho(z^*, z^{**}) = \| z^* - z^{**} \|_K .$$

Переписывая /I6/ в виде

$$y^{(s+1)}_n = g^n(y^{(s)}_n) = (\Lambda^n)^{-1}\left(\Delta y^{n-1} + \tau f^n(y^{(s)}_n)\right),$$

получим:

/29/ $$\| g^n(z^*) - g^n(z^{**}) \|_K \leq \| (\Lambda^n)^{-1} \|_K \cdot \tau \| f^n(z^*) - f^n(z^{**}) \|_K .$$

Как отмечалось выше, Λ^n имеет блочную структуру, и это свойство сохраняется для обратного оператора. Поэтому для матрицы $(\Lambda^n)^{-1}$ в норме, соответствующей K – норме /I4/:

$$\| (\Lambda^n)^{-1} \|_K \leq \max_k \| (\Lambda^n_k)^{-1} \|_n .$$

Однако, оценки /26/, /27/ таковы, что они сохраняются для любых k . Поэтому,

/30/ $$\| (\Lambda^n)^{-1} \|_K \leq \frac{1}{(1 - \tau L_v)(1 - \tau L_w)} .$$

Далее, из условия /4/ после перехода к векторам /6/, /I4/,

следует:

$$\| f^n(y^{n*}) - f^n(y^{n**}) \|_K \le \sum_{m=1}^{K} \| y_m^{n*} - y_m^{n**} \|_n \cdot \sum_{k=1}^{K} L_k^m .$$

Если обозначить

$$L_f = \max_m \sum_{k=1}^{K} L_k^m ,$$

то очевидно

/3I/ $$\| f^n(z^*) - f^n(z^{**}) \|_K \le L_f \| z^* - z^{**} \|_K .$$

Подставляя /30/ и /3I/ в /29/, получим:

$$\| g^n(z^*) - g^n(z^{**}) \|_K \le \frac{\tau L_f}{(1 - \tau L_v)(1 - \tau L_w)} \| z^* - z^{**} \|_K .$$

Следовательно, для выполнения /28/ достаточно потребовать

$$\frac{\tau L_f}{(1 - \tau L_v)(1 - \tau L_w)} = q < 1,$$

которое заведомо выполняется при

/32/ $$\tau < \frac{q}{L_v + L_w + L_f} .$$

Таким образом, вместе с /2I/, /22/, /23/ условия /32/ достаточны для сходимости итерационного процесса /I6/. Отметим, что в /32/ не входят h_r и h_z.

4. Рассмотрим достаточные условия для корректности разностной задачи /I5/. Существование решения следует из существования и ограниченности обратного оператора $(\Lambda^n)^{-1}$. Займемся устойчивостью схемы по начальным условиям, граничным условиям и по правой /нелинейной/ части. Пусть $\bar{y}^n$ есть решение возмущенной задачи

$$\wedge^n \bar{y}^n = \Delta \bar{y}^{n-1} + \tau \bar{f}^n(\bar{y}^n), \quad \bar{y}^o = \bar{u}^o ,$$

причем для $\bar{f}^n$ будем предполагать выполненными условия Липшица /4/ с постоянными $\bar{L}_k^m$. Обозначим разность решений исходной задачи /I5/, /II/ и возмущенной

$$\bar{z}^n = y^n - \bar{y}^n$$

и выпишем задачу для $\bar{z}^n$

$$\wedge^n \bar{z}^n = \Delta \bar{z}^{n-1} + \tau\left[f^n(y^n) - \bar{f}^n(\bar{y}^n)\right], \quad \bar{z}^o = u^o - \bar{u}^o.$$

Перепишем это уравнение:

$$\bar{z}^n = (\wedge^n)^{-1}\left[\Delta \bar{z}^{n-1} + \tau\left(f^n(y^n) - \bar{f}^n(y^n)\right) + \right.$$
$$\left. + \tau\left(\bar{f}^n(y^n) - \bar{f}^n(\bar{y}^n)\right)\right] .$$

Переходя к К - норме и пользуясь /30/, получим:

/33/ $$\|\bar{z}^n\|_K \le \frac{1}{(1 - \tau L_v)(1 - \tau L_w)}\left[\|\Delta \bar{z}^{n-1}\|_K + \tau r_n + \bar{L}_f \|\bar{z}^n\|_K\right],$$

где

$$r_n = \|f^n(y^n) - \bar{f}^n(y^n)\|_K , \quad \bar{L}_f = \max_m \sum_{k=1}^{K} \bar{L}_k^m .$$

Заметим далее, что $\|\Delta \bar{z}^n\|_K = \|\bar{z}^n\|_K$. Неравенство /33/ можно переписать, разрешая относительно нормы $\bar{z}^n$:

$$\|z^n\|_K \le \frac{1}{1 - \tau(\bar{L}_f + L_v + L_w)} \cdot \left[\|\bar{z}^{n-1}\|_K + \tau r_n\right].$$

Подставляя вместо $\|\bar{z}^{n-1}\|_K$ его оценку через $\|\bar{z}^{n-2}\|_K$ и продолжая этот процесс до $n = 0$, получим:

/34/ $$\|\bar{z}^n\|_K \le R^n \|\bar{z}^o\|_K + \tau r_n \sum_{i=o}^{n-1} R^i ,$$

$$R(\tau) = \left[1 - \tau\left(\bar{L}_f + L_v + L_w\right)\right]^{-1}.$$

В [6] показано, что при

$$\tau \le \frac{1}{2}\left(\bar{L}_f + L_v + L_w\right)$$

имеет место оценка

$$R(\tau) \le \exp\left[2\tau\left(\bar{L}_f + L_v + L_w\right)\right].$$

Поэтому, с учетом того, что $\tau N = T$, из /34 следует:

$$\|\bar{z}^n\|_K \le \exp\left[2T\left(\bar{L}_f + L_v + L_w\right)\right]\left[\|\bar{z}^o\|_K + Tr_n\right].$$

Поскольку коэффициент в правой части не зависит от n, получаем выражение устойчивости решения

/36/ $$\max_n \|y^n - \bar{y}^n\|_K \le K_1 \|u^o - \bar{u}^o\|_K + K_2 \max_n \|f^n(y^n) - \bar{f}^n(y^n)\|_K,$$

где значения K_1 и K_2 очевидны.

5. Рассмотрим теперь достаточные условия для сходимости решения разностной краевой задачи /I5/, /II/ к решению дифференциальной задачи /2/, /5/. Спроектируем решение дифференциальной задачи $u_k(r,z,t)$, $k = 1,\dots K$ на сетку и введем вектор $u^n \in Y \subset \mathbb{R}^{K\times(J+1)\times(I+1)}$. Обозначим

$$z^n = y^n - u^n,$$

и для z^n запишем краевую задачу:

/37/ $$\Lambda^n z^n = \Delta z^{n-1} + \tau\left(f^n(y^n) - f^n(u^n)\right) + \tau Q^n,$$

$$z^o = 0,$$

где Q^n – невязка – погрешность аппроксимации системы дифференциальных уравнений, и

$$\|Q^n\|_K = O\left(\tau + \frac{h_r^2}{r} + h_z^2\right).$$

Используя систему доказательства, приведенную выше, при

/38/ $$\tau < \frac{1}{2}\left(L_f + L_v + L_w\right)$$

получим равенство, выражающее сходимость решения разностной задачи:

$$\max_n \|z^n\|_K \leq T \exp\left[2T\left(L_f + L_v + L_w\right)\right] \max_n \|Q^n\|_K =$$

/39/ $$= O\left(\tau + \frac{h_r^2}{2} + h_z^2\right).$$

Поскольку неравенство /39/ не предполагает зависимости между параметрами сетки, т.е. оно справедливо при произвольных τ, h_r и h_z, то мы получили безусловную сходимость в равномерной метрике.

Если сравнить ограничения на τ, h_r и h_z /21/, /22/, /23/, а также /32/ и /38/, полученные в процессе исследования всех этапов численного решения дифференциальной краевой задачи /2/, /5/, то видно, что ограничения на h_r и h_z одни и те же, /21/, /23/, на всех этапах решения. В то же время, достаточные условия для корректности и сходимости нелинейной разностной задачи приводят к более жестким требованиям на τ, чем достаточные условия для устойчивости численного метода решения линеаризированной системы.

Литература.

1. R.H.Farzan, G.Molnárka. Numerical solution of nonlinear parabolic equations with axial symmetry. Numerikal módszerek, ELTE TTK Numerikus és Gépi Matematikai Tanszék, 11 / 1978.

2. Д.Молнарка, Р.Х.Фарзан. О применении неявных разностных схем для решения систем дифференциальных уравнений параболического типа со слабой нелинейностью.
Annales Univ. Sci. Budapest
Sectio Computatorica, /в печати/ .

3. А.А.Самарский. Теория разностных схем. Москва, "Наука", 1977.

4. И.В.Фрязинов. О разностных схемах для уравнения Пуассона в полярной, цилиндрической и сферической системах координат.
Ж. "Вычислительная математика и математическая физика". № 5, №1971.

5. Н.С.Бахвалов. Численные методы. Москва, "Наука", 1973.

6. Д.Молнарка, Р.Х.Фарзан, Л.Фаи. О применении неявной разностной схемы для решения одномерного дифференциального уравнения параболического типа со слабой нелинейностью.
MTA SZTAKI. Közlemények, 21 / 1978.

GENERALIZATION OF THE METHOD OF CONJUGATE GRADIENTS: THE METHOD OF CONJUGATE PAIRS

by

Csaba J. Hegedüs
Central Research Institute for Physics,
Budapest

АННОТАЦИЯ

Метод сопряженных градиентов является эффективным способом для решения линейных систем, имеющих большие разреженные, положительно определенные матрицы. В том случае, когда матрица линейного уравнения Ax=b является любой разреженной матрицей, можно применить метод сопряженных градиентов для нормального уравнения $A^TAx = A^Tb$. Но теперь спектральное число обусловленности возводится в квадрат поэтому сходимость метода замедляется, особенно в случаях плохо обусловленных матриц.

Чтобы обойти упомянутые проблемы возможны следующие подходы;

/I/ Уменьшение спектрального числа обсусловленности с помощью подходящей нормировки матрицы;

/II/ Расширение метода сопряженных градиентов для любых матриц вместо увеличения индекса обусловленности.

Настоящая статья является первым сообщением об успешной реализации второго подхода. Определяются обобщенные рекуррентные формулы A-ортогональных пар векторов /сопряженных парах/.

CONTENT

The method of conjugate gradients is an effective method for solving large sparse sets of linear equations when the coefficient matrix A is positive /or negative/ definite. In cases when having a linear equation Ax=b with an arbitrary sparse matrix A, one can apply the method by changing to the normal equation $A^TAx=A^Tb$. Now the condition number is squared that leads to a slower rate of convergence, especially in cases of ill-conditioned matrices, i.e. matrices with large condition numbers. For curing the problems, one may try two approaches:

i/ Decrease the condition number by equilibration;

ii/ Try to extend the method of conjugate gradients for arbitrary matrices such that the condition number is not increased.

This paper is a first report on successful realization of the second goal. Generalized recursive formulae are obtained for generating A-orthogonal pairs of vectors or as they are called shortly: conjugate pairs. A detailed analysis for numerical applications will be the topic of a subsequent paper.

1. INTRODUCTION

The method of conjugate gradients was developed by Hestenes and Stiefel in 1952 [1]. Applicable to systems of linear equations with symmetric positive definite matrices, it is not recommended as a direct method in contrast to Gaussian elimination because of the amount of numerical operations involved. However, as an iterational method it is competitive with other methods such as Chebyshev iteration or SOR methods when large sparse sets of linear equations need to be solved [2].

If $\varkappa$ denotes the condition number of the matrix in the linear equation then the convergence of the method is characterized by the inequality

$$\|x_k - h\|_2 \leq C\{(\sqrt{\varkappa} - 1)/(\sqrt{\varkappa} + 1)\}^k \qquad (1)$$

in the kth step, where C is some constant and $\|x_k - h\|_2$ is the Euclidean distance of x_k, the kth approximation, from the solution h [3],[4].

The method can also be used to solve the linear equation $Ax=b$ with a general coefficient matrix by solving the normal equation $A^TAx=A^Tb$. But, from computational point of view it is known that the condition number of the coefficient matrix is now squared [5], hence making an ill-conditioned system even more ill-conditioned that needs a larger amount of iterations according to the theory and computational experiences.

The most general forms of the method of conjugate gradients can be found in a paper by Hestenes [7]. From those generalizations a more stable version for the case of an arbitrary coefficient matrix was suggested again by Clasen [6]. This method minimizes the residuals directly in each step by using another metric in the inner products, but requires twice as many operations than the popular method.

The present paper makes a different approach by dropping the minimization of a quadratic form and recursive formulae are derived for generating A-orthogonal pairs of vectors for an arbitrary matrix. For brevity, the term 'conjugate pairs' is used for these vectors throughout the paper. The derivation is based on introducing appropriate projection operations and the solution of a linear system is expressed in terms of these vectors.

The new method in its present form needs further analysis before any recommendation can be made for numerical purposes. Applying some results of singular value decomposition, there are indications that for certain starting vectors all attractive features of the conjugate gradients are preserved and the condition number is not squared at the same time. These results will be detailed in a subsequent paper.

2. GENERAL FORMULATION FOR CONJUGATE PAIRS

Let $A \in C^{m,n}$, $v_j \in C^m$, $u_j \in C^n$ and assume the elements of the set $\{v_j, u_j\}_{j=1}^{i}$ satisfy the relations:

$$v_j^H A u_k = \lambda_k \delta_{jk}, \quad j,k=1,2,\ldots,i, \quad \lambda_k \neq 0, \tag{2}$$

where δ_{jk} is the Kronecker delta and H stands for the conjugate transpose. The vectors of the set $\{v_j, u_j\}_{j=1}^{i}$ are said to be a system of conjugate pairs with respect to matrix A. Moreover, introduce the projectors

$$P_i^r = I_n - \sum_{j=1}^{i} u_j v_j^H A / v_j^H A u_j, \tag{3}$$

$$P_i^\ell = I_m - \sum_{j=1}^{i} A u_j v_j^H / v_j^H A u_j, \tag{4}$$

where I_n is the unit matrix of order n and the superscripts r and ℓ refer to the relative positions of matrix A in the numerator, viz. right and left. One can easily check the following properties of these matrices:

$$P_i^r P_i^r = P_i^r, \tag{5}$$

$$P_i^r u_j = 0, \quad j \leq i, \tag{6}$$

$$v_j^H A P_i^r = 0, \quad j \leq i, \tag{7}$$

and

$$P_i^\ell P_i^\ell = P_i^\ell, \tag{8}$$

$$v_j^H P_i^\ell = 0, \quad j \leq i, \tag{9}$$

$$P_i^\ell A u_j = 0, \quad j \leq i. \tag{10}$$

The projectors have the ranks:

$$\varrho(P_i^r) = Tr(P_i^r) = n-i \tag{11}$$

and

$$\varrho(P_i^\ell) = Tr(P_i) = m-i, \tag{12}$$

indicating that the systems $\{u_j\}_{j=1}^i, \{v_j\}_{j=1}^i, \{Au_j\}_{j=1}^i$ and $\{v_j^H A\}_{j=1}^i$ have linearly independent vectors that follows from standard matrix theory.

With the aid of projectors (3) and (4), one can generate a new pair into the set $\{v_j, u_j\}_{j=1}^i$. Let $q_{i+1} \in C^n$ and $r_{i+1} \in C^m$ be two vectors such that

$$r_{i+1}^H P_i A P_i^r q_{i+1} \neq 0. \tag{13}$$

Then the vectors

$$u_{i+1} = P_i^r q_{i+1} \tag{14}$$

and

$$v_{i+1}^H = r_{i+1}^H P_i \tag{15}$$

form a new conjugate pair - as it can be checked from relations (7) and (10). In the case of $P_i^\ell A P_i^r \neq 0$, one can surely find two vectors that fulfil condition (13).

In the other case when for an $i=\varrho$ the relation $P_\varrho^\ell A P_\varrho^r = 0$ holds, that is,

$$A = \sum_{j=1}^{\varrho} Au_j v_j^H A / v_j^H Au_j, \tag{16}$$

a rank factorization of matrix A is attained with the pair of vectors Au_j and $v_j^H A$, $j=1,2,\ldots,\varrho$. These vectors are linear-

ly independent indicating the rank of matrix A:

$$\varrho(A) = \varrho. \tag{17}$$

Hence, a system of conjugate pairs with respect to matrix A may contain no more independent pairs of elements than the rank of matrix A. If this is the case, then the system is said to be complete with respect to A.

Now let (17) hold and $\{v_j, u_j\}_{j=1}^{\varrho}$ form a complete system of conjugate pairs. Then the linear equation

$$Ax = b \tag{18}$$

is consistent if $P_{\varrho}^{\ell} b = 0$ and the solutions are obtained as

$$x = \sum_{j=1}^{\varrho} u_j v_j^H b / v_j^H A u_j + P^r t, \tag{19}$$

where t is an arbitrary vector in C^n. In fact, condition $P_{\varrho}^{\ell} b=0$ is necessary as $P_{\varrho}^{\ell} A=0$ holds equivalently to (16), yielding $P_{\varrho}^{\ell} Ax = P_{\varrho}^{\ell} b = 0$. On the other hand, it is sufficient because then b is an element of the linear subspace spanned by the column vectors of A. From $P_{\varrho}^{\ell} b=0$ and (4) one gets

$$b = A \sum_{j=1}^{\varrho} u_j v_j^H b / v_j^H A u_j \tag{20}$$

showing a particular solution to (18). The general solution is given by (19) as $P_{\varrho}^{r} t$ is the general solution to the homogeneous equation $Ax=0$.

Introduce the matrices

$$V = [v_1, v_2, \ldots, v_{\varrho}], \qquad U = [u_1, u_2, \ldots, u_{\varrho}] \tag{21}$$

then (16) has the following form

$$A = AU(V^H AU)^{-1} V^H A, \tag{22}$$

where V^HAU is a diagonal matrix. The matrix $X=U(V^HAU)^{-1}V^H$ is a (1,2)-generalized inverse to A since together with (22), XAX=X also holds. According to the theory of generalized inverses [8], the first term in (19) does not necessarily yield a least squares solution to (18). This is in contrast to the result one can get for the conjugate gradient method in the case of positive semidefinite operators [4].

3. RECURSIVE GENERATION OF CONJUGATE PAIRS

It will be shown in the following that conjugate pairs can be generated in a recursive manner such that two auxiliary vectors and a conjugate vector pair are needed only for the generation of the next pair.

Let

$$u_{i+1} = P_i^r q_{i+1}, \tag{23}$$

and

$$v_{i+1}^H = r_{i+1}^H P_i^\ell \tag{24}$$

as it was demanded before and let the auxiliary vectors q_{i+1} and r_{i+1} be chosen such that

$$q_{i+1}^H = q_i^H P_i^r \tag{25}$$

and

$$r_{i+1} = P_i^\ell r_i. \tag{26}$$

One can show that it is enough to take the last ith term in the sums of (3) and (4) when the projectors P_i^r and P_i^ℓ are applied in (23)-(26). From the properties of the projectors (see (5)-(10)), one has

$$q_k^H u_j = q_{k-1}^H P_{k-1}^r u_j = 0, \qquad j < k, \tag{27}$$

$$v_j^H r_k = v_j^H P_{k-1}^{\ell} r_{k-1} = 0, \qquad j < k. \tag{28}$$

Hence (25) and (26) take the simpler forms:

$$q_{i+1}^H = q_i^H (I_n - u_i v_i^H A / v_i^H A u_i), \tag{29}$$

$$r_{i+1} = (I_m - A u_i v_i^H / v_i^H A u_i) r_i. \tag{30}$$

In order to justify the above statement for (23) and (24), one observes that the vectors q_i and r_i can be expanded as

$$q_i = \sum_{j=1}^{i} \alpha_{ij} u_j \tag{31}$$

and

$$r_i = \sum_{j=1}^{i} \beta_{ij} v_j^H. \tag{32}$$

Now, if $i < k$ then an inspection of formulae (27) and (28) yields the following relations:

$$q_k^H q_i = \sum_{j=1}^{i} \alpha_{ij} q_k^H u_j = 0, \tag{33}$$

$$r_i^H r_k = \sum_{j=1}^{i} \beta_{ij} v_j^H r_k = 0. \tag{34}$$

For $k < i$, one gets the same relations showing that the vectors $q_1, q_2, \ldots$ and $r_1, r_2, \ldots$ form orthogonal systems. Thus one concludes from (29) and (30) that the following relations hold:

$$v_i^H A q_k = 0, \quad \text{if} \quad k \neq i \quad \text{or} \quad k \neq i+1, \tag{35}$$

$$r_k^H A u_i = 0, \quad \text{if} \quad k \neq i \quad \text{or} \quad k \neq i+1. \tag{36}$$

Putting these into (23) and (24) and using the actual forms for the projectors P_i^r and P_i^ℓ, one gets:

$$u_{i+1} = \Big(I_n - \sum_{j=1}^{i} u_j v_j^H A / v_j^H A u_j\Big) q_{i+1} \tag{37a}$$

$$= \left(I_n - u_i v_i^H A / v_i^H A u_i\right) q_{i+1},$$

$$v_{i+1}^H = r_{i+1}^H \Big(I_m - \sum_{j=1}^{i} A u_j v_j^H / v_j^H A u_j\Big) \tag{38a}$$

$$= r_{i+1}^H \left(I_m - A u_i v_i^H / v_i^H A u_i\right).$$

Note that alternative forms can be obtained for u_{i+1} and v_{i+1} if one substitutes $v_i^H A$ and $A u_i$ from (29) and (30):

$$u_{i+1} = q_{i+1} + u_i \|q_{i+1}\|_2^2 / \|q_i\|_2^2, \tag{37b}$$

$$v_{i+1} = r_{i+1} + v_i \|r_{i+1}\|_2^2 / \|r_i\|_2^2, \tag{38b}$$

where the relations $q_j^H u_j = \|q_j\|_2^2$ and $v_j^H r_j = \|r_j\|_2^2$ were used that can be seen from (37a) and (38a) if one multiplies by q_{i+1}^H and r_{i+1}^H respectively and takes (27) and (28) into account.

The formulae (27),(28),(37a,b),(38a,b) coincides with those of the conjugate gradient method if matrix A is symmetric and $q_1 = r_1 = v_1 = u_1$.

References:

[1] M.R. H e s t e n e s and E. S t i e f e l: Methods of conjugate gradients for solving linear systems. NBS J. Res. 49, 409-436 (1952).

[2] J.K. R e i d: On the method of conjugate gradients for the solution of large sparse systems of linear equations. In: Large Sparse Sets of Linear Equations (Ed. J.K. R e i d) Acad. Press, London, 1971. pp. 231-254.

[3] J.W. D a n i e l: The conjugate gradient method for linear and nonlinear operator equations. SIAM J. Numer. Anal. 4, 10-26 (1967).

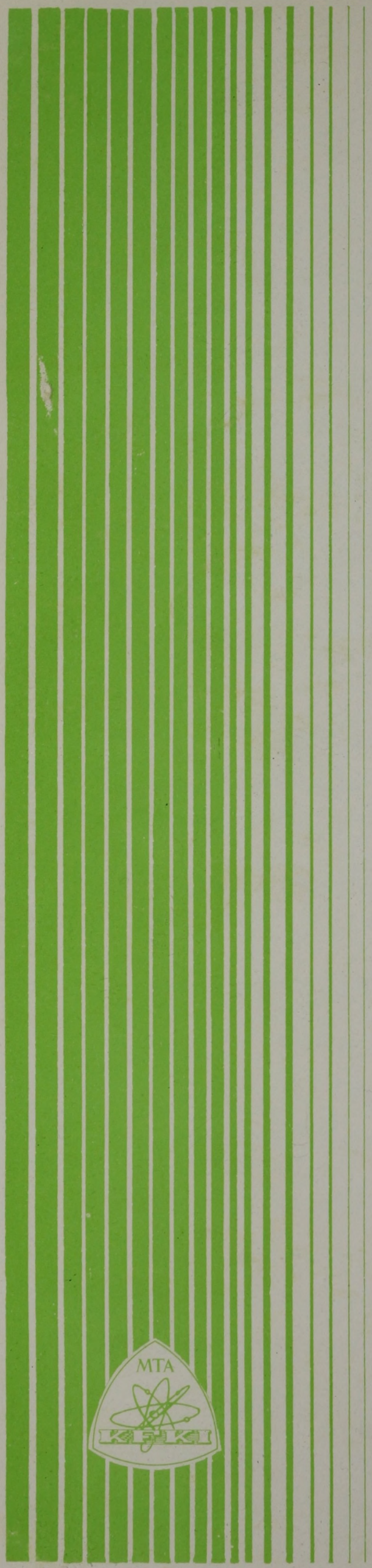
[4] W.J. K a m m e r e r and M.Z. N a s h e d: On the convergence of the conjugate methods for singular operator equations. SIAM J. Numer. Anal. 9, 165-181 (1972).

[5] G.W. S t e w a r t: Introduction to Matrix Computations. Acad. Press, New York, 1973. p. 223.

[6] R.J. C l a s e n: A note on the use of the conjugate gradient method in the solution of a large system of sparse equations. The Computer Journal. 2o, 185-186 (1977).

[7] M.R. H e s t e n e s: The conjugate gradient method for solving linear systems. Proc. Symp. Appl. Math. 6, 83-102 (1956).

[8] C.R. R a o and S.K. M i t r a: Generalized Inverse of Matrices and its Applications. John Wiley, New York, 1971.



Felelős kiadó: Sándory Mihály
Szakmai lektor: Németh Géza
Nyelvi lektor: Harvey Shenker
Példányszám: 420 Törzsszám:79-868
Készült a KFKI sokszorosító üzemében
Budapest, 1979. november hó